

Applicants: Rachel E. Meyers, et al.
 Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
 Attorney/Agent: Kerri Pollard Schray
 Docket No.: MPI00-079P1RCP2CN1M
 Sheet 1 of 43

1/43

Input file Fbh32142FL.seq; Output File 32142.trans
 Sequence length 2660

CCTTTNTNRCCACGCGTCCGAGAGCGCCCCGCAGTCTTCGGAAAGCGTTGGGTAGGCC	ATG GCT GCG ACG	4
CGT GCA GGG CCC CGC GCC CGC GAG ATC TTC ACC TCG CTG GAG TAC GGA CCG GTG CCG GAG		12
S H A C A L A W L D T Q D R C L G H Y V		24
AGC CAC GCA TGC GCA CTG GCC TGG CTG GAC ACC CAG GAC CGG TGC TTG GGC CAC TAT GTG		72
N G K W L K P E H R N S V P C Q D P I T		44
AAT GGG AAG TGG TTA AAG CCT GAA CAC AGA AAT TCA GTG CCT TGC CAG GAT CCC ATC ACA		132
G E N L A S C L Q A Q A E D V A A A V E		84
GGA GAG AAC TTG GCC AGT TGC CTG CAG GCA CAG GCC GAG GAT GTG GCT GCA GCC GTG GAG		252
A A R M A F K G W S A H P G V V R A Q H		104
GCA GCC AGG ATG GCA TTT AAG GGC TGG AGT GCG CAC CCC GGC GTC GTC CGG GCC CAG CAC		312
L T R L A E V I Q K H Q R L L W T L E S		124
CTG ACC AGG CTG GCC GAG GTG ATC CAG AAG CAC CAG CGG CTG CTG TGG ACC CTG GAA TCC		372
L V T G R A V R E V R D G D V Q L A Q Q		144
CTG GTG ACT GGG CGG GCT GTT CGA GAG GTT CGA GAC GGG GAC GTC CAG CTG GCC CAG CAG		432
L L H Y H A I Q A S T Q E E A L A G W E		164
CTG CTC CAC TAC CAT GCA ATC CAG GCA TCC ACC CAG GAG GAG GCA CTG GCA GGC TGG GAG		492
P M G V I G L I L P P T F S F L E M M W		184
CCC ATG GGA GTA ATT GGC CTC ATC CTG CCA CCC ACA TTC TCC TTT GAG ATG ATG TGG		552
R I C P A L A V G C T V V A L V P P A S		204
AGG ATT TGC CCT GCC CTG GCT GTG GGC TGC ACC GTG GTG GCC CTC GTG CCC CCG GCC TCC		612
P A P L L A Q L A G E L G P F P G I L		224
CCG GCG CCC CTC CTC CTG GCC CAG CTG GCG GGG GAG CTG GGC CCC TTC CCG GGA ATC CTG		672
N V V S G P A S L V P I L A S Q P G I R		244
AAT GTC GTC AGT GGC CCT GCG TCC CTG GTG CCC ATC CTG GCC TCC CAG CCT GGA ATC CGG		732
K V A F C G A P E E G R A L R R S L A G		264
AAG GTG GCC TTC TGC GGA GCC CCG GAG GAA GGG CGT GCC CTT CGA CGG AGC CTG GCG GGA		792
E C A E L G L A L G T E S L L L L T D T		284
GAG TGT GCG GAG CTG GGC CTG GCG CTG GGG ACG GAG TCG CTG CTG CTG ACN GAC ACG		852

Fig. 1A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
 Sheet 2 of 43

2/43

A	D	V	D	S	A	V	E	G	V	V	D	A	A	W	S	D	R	G	P	304
GCG	GAC	GTA	GAC	TCG	GCC	GTG	GAG	GGT	GTC	GTG	GAC	GCC	GCC	TGG	TCC	GAC	CGC	GGC	CCG	912
G	G	L	R	L	L	I	Q	E	S	V	W	D	E	A	M	R	R	L	Q	324
GGT	GGC	CTC	AGG	CTC	CTC	ATC	CAG	GAG	TCT	GTG	TGG	GAT	GAA	GCC	ATG	AGA	CGG	CTG	CAG	972
E	R	M	G	R	L	R	S	G	R	G	L	D	G	A	V	D	M	G	A	344
GAG	CGG	ATG	GGG	CGG	CTT	CGG	AGT	GGC	CGA	GGG	CTG	GAT	GGG	GCC	GTG	GAC	ATG	GGG	GCC	1032
R	G	A	A	A	C	D	L	V	Q	R	F	V	R	E	A	Q	S	Q	G	364
CGG	GGG	GCT	GCC	GCA	TGT	GAC	CTG	GTC	CAG	CGC	TTT	GTG	CGT	GAG	GCC	CAG	AGC	CAG	GGT	1092
A	Q	V	F	Q	A	G	D	V	P	S	E	R	P	F	Y	P	P	T	L	384
GCA	CAG	GTG	TTC	CAG	GCT	GGT	GAT	GTG	CCT	TCG	GAA	CGC	CCA	TTC	TAT	CCC	CCA	ACC	TTG	1152
V	S	N	L	P	P	A	S	P	C	A	Q	V	E	V	P	W	P	V	V	404
GTC	TCC	AAC	CTG	CCC	CCA	GCC	TCC	CCA	TGT	GCC	CAG	GTG	GAG	GTG	CCG	TGG	CCT	GTG	GTC	1212
V	A	S	P	F	R	T	A	K	E	A	L	L	V	A	N	G	T	P	R	424
GTG	GCC	TCC	CCC	TTC	CGC	ACA	GCC	AAG	GAG	GCA	CTG	TTG	GTG	GCC	AAC	GGG	ACG	CCC	CGC	1272
G	G	S	A	S	V	W	S	E	R	L	G	Q	A	L	E	L	G	Y	G	444
GGG	GGC	AGC	GCC	AGT	GTG	TGG	AGC	GAG	AGG	CTG	GGG	CAG	GCG	CTG	GAG	CTG	GGC	TAT	GGG	1332
L	Q	V	G	T	V	W	I	N	A	H	G	L	R	D	P	S	V	P	T	464
CTC	CAG	GTG	GGC	ACT	GTC	TGG	ATC	AAC	GCC	CAC	GGC	CTC	AGA	GAC	CCT	TCG	GTG	CCC	ACA	1392
G	G	C	K	E	S	G	C	S	W	H	G	G	P	D	G	L	Y	E	Y	484
GCG	GGC	TGC	AAG	GAG	AGT	GGG	TGT	TCC	TGG	CAC	GGG	GGC	CCA	GAC	GGG	CTG	TAT	GAG	TAT	1452
L	R	P	S	G	T	P	A	R	L	S	C	L	S	K	N	L	N	Y	D	504
CTG	CGG	CCC	TCA	GGG	ACC	CCT	GCC	CGG	CTG	TCC	TGC	CTC	TCC	AAG	AAC	CTG	AAC	TAT	GAC	1512
T	F	G	L	A	V	P	S	T	L	P	A	G	P	E	I	G	P	S	P	524
ACC	TTT	GGC	CTC	GCT	GTG	CCC	TCA	ACC	CTG	CCG	GCT	GGG	CCT	GAA	ATA	GGG	CCC	AGC	CCA	1572
A	P	P	Y	G	L	F	V	G	G	R	F	Q	A	P	G	A	R	S	S	544
GCA	CCC	CCC	TAT	GGG	CTC	TTC	GTG	GGG	GGC	CGT	TTC	CAG	GCT	CCT	GGG	GCC	CGA	AGC	TCC	1632
R	P	I	R	D	S	S	G	N	L	H	G	Y	V	A	E	G	G	A	K	564
AGG	CCC	ATC	CGG	GAT	TCG	TCT	GGC	AAT	CTC	CAT	GGC	TAC	GTG	GCT	GAG	GGT	GGA	GCC	AAG	1692
D	I	R	G	A	V	E	A	A	H	Q	A	F	P	G	W	A	G	Q	S	584
GAC	ATC	CGA	GGT	GCT	GTG	GAG	GCC	GCT	CAC	CAG	GCT	TTC	CCT	GGC	TGG	GCG	GGC	CAG	TCC	1752
P	G	A	R	A	A	L	L	W	A	L	A	A	A	L	E	R	R	K	S	604
CCA	GGA	GCC	CGG	GCA	GCC	CTG	CTG	TGG	GCC	CTG	GCG	GCT	GCA	CTG	GAG	CGC	CGG	AAG	TCT	1812

Fig. 1B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 3 of 43

3/43

T L A S R L E R Q G A E L K A A E A E V 624
ACC CTG GCC TCA AGG CTG GAG AGG CAG GGA GCG GAG CTC AAG GCT GCG GAG GCG GAG GTG 1872

E L S A R R L R A W G A R V Q A Q G H T 644
GAG CTG AGC GCA AGA CGA CTT CGG GCG TGG GGG GCC CGG GTG CAG GCC CAA GGC CAC ACC 1932

L Q V A G L R G P V L R L R E P L G V L 664
CTG CAG GTA GCC GGG CTG AGA GGC CCT GTG CTG CGC CTG CGG GAG CCG CTG GGT GTG CTG 1992

A V V C P D E W P L L A F V S L L A P A 684
GCT GTG GTG TGT CCG GAC GAG TGG CCC CTG CTT GCC TTC GTG TCC CTG CTG GCT CCC GCC 2052

L A Y G N T V V M V P S A A C P L L A L 704
CTG GCC TAC GGC AAC ACT GTG GTC ATG GTG CCC AGT GCG GCC TGT CCT CTG CTG GCC CTG 2112

E V C Q D M A T V F P A G L A N V V T G 724
GAG GTC TGC CAG GAC ATG GCC ACC GTG TTC CCA GCA GGC CTG GCC AAC GTG GTG ACA GGA 2172

D R D H L T R C L A L H Q D V Q A M W Y 744
GAC CGG GAC CAT CTG ACC CGC TGC CTG GCC TTG CAC CAA GAC GTC CAG GCC ATG TGG TAT 2232

F G S A Q G S Q F V E W A S A G N L K P 764
TTC GGA TCA GCC CAG GGT TCC CAG TTT GTC GAG TGG GCC TCG GCA GGA AAC CTC AAA CCG 2292

V W A S R G C P R A W D Q E A E G A G P 784
GTG TGG GCG AGC AGG GGC TGC CCG CGG GCC TGG GAC CAG GAG GCC GAG GGG GCA GGC CCA 2352

E L G L R V A R T K A L W L P M G D * 803
GAG CTG GGG CTG CGA GTG GCG CGG ACC AAG GCC CTG TGG CTG CCT ATG GGG GAC TGA 2409

TGCCTGAGGCCACCTACTGCATTTGGACACCTCACACCAAGGGGAGATGCACCCACAGACACCTGGACTTCCCC
TTCTGGTTCTGTGTCTCCAATAACTCTGACCAACCTAAAAAAAAAAAAARWARMAACTTC
TGGCAGATATGAGGCTTTTTCTTTTTTT

Fig. 1C

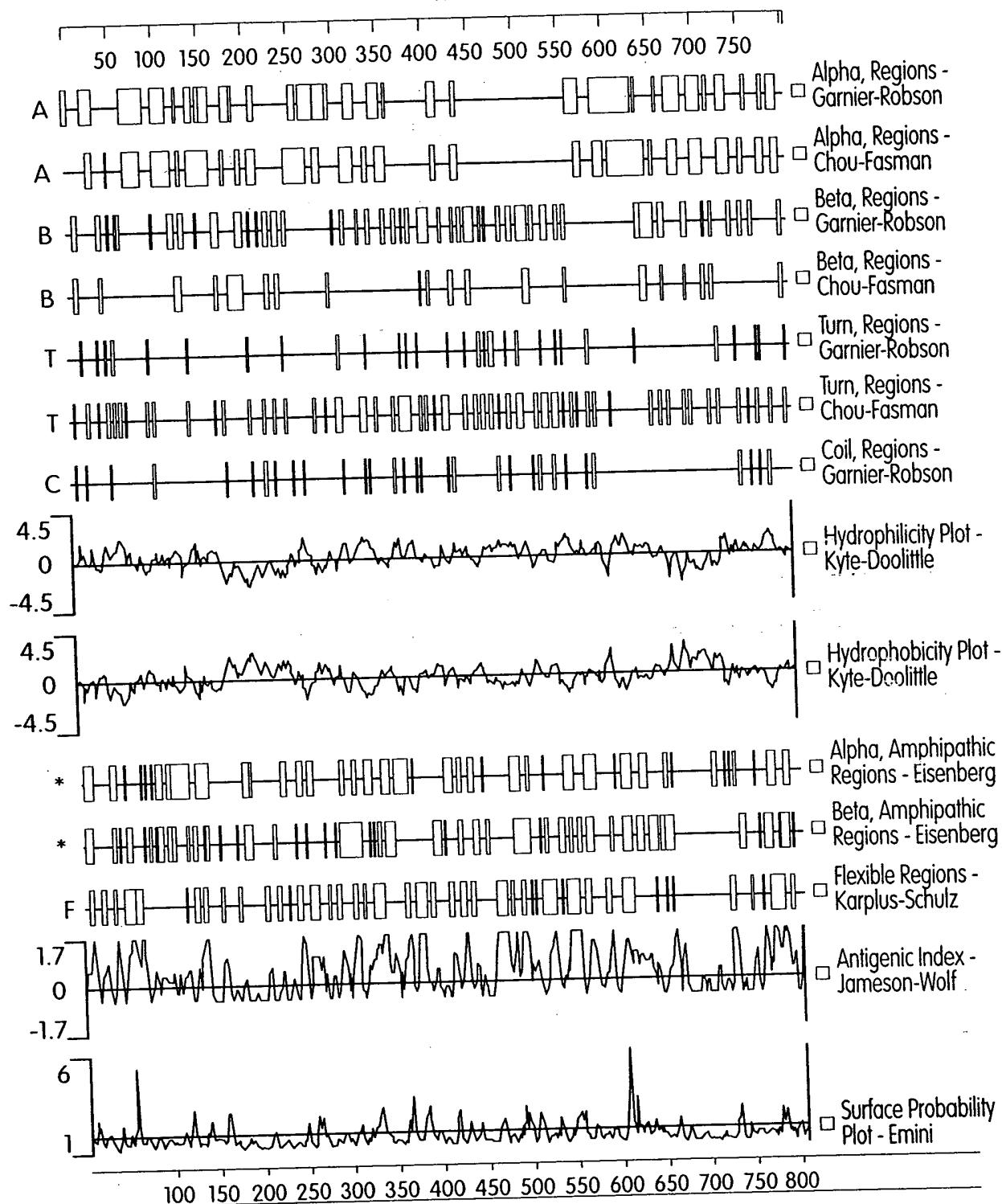


Fig. 2

5/43

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
159	175	ins-->out	0.1

>32142
MAATRAGPRAREIFTSLEYGPVPESHACALAWLDTQDRCLGHYVNGKWLKPEHRNSVPCQ
DPITGENLASCLQAQAEDVAAAVEAARMAFKGWSAHPGVRAQHLTRLAEVIQKHQRLLW
TLESLVTGRAVREVRDGDVQLAQQLLHYHAIQASTQEEALAGWEPMGVIGLILPPTFSFL
EMMWRICPALAVGCTVALVPPASPAPLLAQLAGELGPFPGLNVVSGPASLVPILASQ
PGIRKVAFCGAPEEGRALRRSLAGECAELGLALGTESLLLLTDTADVDSAVEGVVDAAWS
DRGPGGLRLLIQESVWDEAMRLIQERMGRRLSGRGLDGAVDMGARGAACDLVQRFVREA
QSQGAQVFQAGDVPSERPFYPPTLVSNLPASPACAQVEVPWPVVVASPFRTAKEALLVAN
GTPRGGSASVWSERLGQALELGQVGTWINAHGLRDPSPVPTGGCKESGCSWHGGPDG
LYEYLRPSTGTPARLSCLSKNLNYDTFGLAVPSTLPAGPEIGPSPAPPYGLFVGGRFQAPG
ARSSRPIRDSSGNLHGYYAEGGAKDIRGAVEAAHQAFPGWAGQSPGARAALLWALAAALE
RRKSTLASRLERQGAELKAAEAEVELSARRLRAWGARVQAQGHTLQVAGLRGPVLRLREP
LGVLAVVCPDEWPLLAFFSLLAPALAYGNTVVMVPSAACPLLALLEVQCDMATVFPAGLAN
VVTGDRDHLTRCLALHQDVQAMWYFGSAQGSQFVEWASAGNLKPVWASRGCPRAWDQEAE
GAGPELGLRVARTKALWLPMD

Fig. 3

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 6 of 43

6/43

Protein Family / Domain Matches, HMMer Version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.0/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.9519.seq

Query: 32142

Scores for sequence family classification (score includes all domains):				
Model	Description	Score	E-value	N
<u>aldedh</u>	Aldehyde dehydrogenase family	149.8	4.7e-41	1

Parsed for domains:

Model	Domain	seq-f, seq-t	hmm-f	hmm-t	score	E-value
aldedh	1/1	47	494 ..	1 492 []	149.8	4.7e-41

Alignments of top-scoring domains:

aldedh	domain 1 of 1, from 47 to 494: score 149.8, E = 4.7e-41
	*->ewvdsasgktfevvNPankgevigrvpeataeDvdaAVkAAkeAfks
	+w +++ + +++ +P + ge +++ +a+aeDv aAV AA+ Afk+
32142	47 KWLWPEHRNSVPCQDPIT-GENLASCLQAQaedaVAAVEAARMAFKG 92
	GpwWakvpaseRariLrkladlieeredeLaaletlDlGKplaeAkgDte
	W++ p Ra+ L +la+ i+ ++ +L le+l +G ++et+ + +
32142	93 ---WSAHPGVVRQHLTRLAEVIQKHQRLLWTLSVTGRAVREVRDG-D 138
	vgraideiryyagwarklmgerrvipslatdgdeelnytrrePlGVvgvI
	v+ a + +++y a +a+ t+ e ++ +eP GV+g I
32142	139 VQLAQQLLHYHAIQAS-----TQ---EEALAGWEPMGVIGLI 172
	SPWNFP111alwklaPAAgNTVV1KPSEqTPlt..all1aelieeaGa
	P F +1 ++w ++pALA G+TVV + P+++ llla l e G
32142	172 LPPTFSFLEMMWRICPALAVGCTVV---ALVPPASpaPLLLAQLAGELG- 218
	n1lPkGVvnvvPgfGaevGqaLlshpdidkisFTGStEVgklimeAAak
	+G +nnv G +a+ + L+s+p+i+k++F G +e G+ + ++ A+
32142	219 --PFPGILNVVSG-PASLVPILASQPGIRKVAFCGAPEEGRALRRSLAGE 265
	n1kkVtLELGGKsPvIVfdDADLdkAverivfgaFgnaGQvCiApsR11v
	+ L LG s d AD d Ave+v+a G ++ R11+
32142	266 -CAELGLALGTESLLLTDTADVDSAVEGVVDAAWSDRG---PGGLRLLI 311
	hesiydeFveklkervkklkliGdpldsdt niyGPlIseqqfdrvlsyIe
	+es+ de + +1+er+ +1+ G +ld + + G+ + ++ d v + ++
32142	312 QESVWDEAMRRLQERMGLR-SGRGLDGAVDM-GAR-GAACDLVQRFVR 358
	dgkeeGAvlcGGerdeskeylggGyyvqPTiftdVtpdMkImkEEIFGP
	+++++GA+v + G ++ + + ++ PT+++++ p +++++ E+ P
32142	359 EAQSQGAQVFQAGDVPSE---RP---FYPPTLVSNLPPASPACAQVEVPWP 402
	VlpiikfklddEAIelaNdteYGLAayvFTkdilarafrvakaleaGiVw
	V++ f++ EA+ aN t+ G +a+v+++ l a +l++G+Vw
32142	403 VVVASFRTAKEALLVANGTPRGGSASVWSER-LGQALELGYGLQVGTWV 451
	vNDvcvhahaepqlPFGGvHqSSGiGrehgGkygleeYteiKtVtirl<-*
	+N ++ +p++P GG K+ SG + ++ G+gl eY++ + rl
32142	452 IN--AHGLRDPSPVTGGCKE-SGCCSWHG-GPDGLYELYLRPSGTPARL 494

Fig. 4

ProDom Matches

ProdomId	Start	End	Description	Score
View Prodom 135 <input type="button" value="Boxer ▼"/>			p99.2 (229) DHAL(10) DHAB(10) DHAM(7) // DEHYDROGENASE OXIDOREDUCTASE ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE	280
Showing match <input type="button" value="▼"/> <input type="button" value="Go!"/>	101	770		

View Prodom 135

>135 p99.2 (229) DHAL(10) DHAB(10) DHAM(7) // DEHYDROGENASE OXIDOREDUCTASE
ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE
Length = 494

Score = 280 (103.6 bits), Expect = 7.8e-22, P = 7.8e-22
Identities = 87/289 (30%), Positives = 142/289 (49%)

Query: 216 ELGPFPGILNVVSG--PASLVPILASQPGIRKVAFCGAPEEGRALRRSXXXXXXXXXXXX 273
E G PG++Nvv+G A + L S P I K++F G+ E G+A+ ++

Sbjct: 194 EAGLPPGVINVVTGFGGAEVGEALVSHPDIDKISFTGSTEVGKAIMAAEKNLKPVTLE 253

Query: 274 XXXX--XXXXDTADVDSAVEGVVDAAWSDRGP---GGLRLLIQESVWDEAMRLQERMG 328
D D+D AVE VV A+ + G R+ +QES++DE + +L ER+

Sbjct: 254 LGGKNPVIVFEDADDLDKAVESVVFAGFFNSGQVCTAASRIFVQESIYDEFVEKLVERVK 313

Query: 329 RL-RSGRG--LDGAVDMGAR-GAACCDLVQRFVREAQSQGAQVFQAGD---VPSERPFY- 380
+L + G LD DMG + +Q ++ EA+++GA++ G+ E ++

Sbjct: 314 KLLKVGEDDPLDPDTDMGPLINEEQYEKIQSYYIEEAKAEGAKLVCGGERRKAGDEGGYFI 373

Query: 381 PP TLVSNLPPASP CQA QV E VP W P V V V A S P F R T - A K E A L L V A N G T P R G G S A S V W S E R L G O A L 439
PT+++++ Q E+ PV+ F+ EA+ +AN T G +A V++ + +A

Sbjct: 374 QPTILTDVTEDMRIMQEEIFGPVLPVIKFKDLD E A I E L A N D T E Y G L A A G V F T R D I E R A Q 433

Query: 440 ELGYGLQVGTWINA---HGLRDPSVPTGGCKESGCSWH-GGPDGLYEF 484
+ L+ GTW+N H + P GG K+SG GG GL EY

Sbjct: 434 RVAERLEAGTVWWNDNIYHVSAEAQAPFGGYKQSGIGGREGGKYGLEY 482

Score = 262 (97.3 bits), Expect = 8.2e-20, P = 8.2e-20
Identities = 86/301 (28%), Positives = 140/301 (46%)

Query: 101 RAQHLTRLAEVIQKHQRLLWTLESVTGRAVREVRDGDVQLAQQLLHYHA----- 150
RA+ L +LA++++++ L LE+L TG+ + E + +V A L Y+A

Sbjct: 61 RARILRKLA DLLEENKDE LA LETLETGKPLAEAKVAEV VARAVDYLRYAGMAE KLM GEE 120

Query: 151 -IQASTQEE----ALAGWEPMGVIGLILPPTFSFLEMMWRICPALAVGVTXX---XXXXX 202

I S E + EP+GV+ I P F + +W+I PALA G T

Sbjct: 121 TIPTSLSESPGSMSYTMR EPLGVVAAITPNFPLMMAVWKIA P ALA AGNTVVLKPSEQTP 180

Query: 203 XXXXXXXXXXXXXXXXGELGPFPGILNVVSG--PASLVPILASQPGIRKVAFCGAPEEGRALRR 260
E G PG++Nvv+G A + L S P I K++F G+ E G+A+ +

Sbjct: 181 LTALLLAELIKEAEAGLPPGVINVVTGFGGAEVGEALVSHPDIDKISFTGSTEVGKAIMK 240

Query: 261 XXXXXXXXXXXXXXXX--XXXXDTADVDSAVEGVVDAAWSDPGP---GGLRLLIQESV 315
+ D D+D AVE VV A+ + G R+ +QES+

Sbjct: 241 AAAEKNLKPVTLELGKNPVIVFEDADDLDKAVESVVFAGFFNSGQVCTAASRIFVQESI 300

Query: 316 WDEAMRLQERMGRL-RSGRG--LDGAVDMGAR-GAACCDLVQRFVREAQSQGAQVFQAG 371
+DE + +L ER+ +L + G LD DMG + +Q ++ EA+++GA++ G

Sbjct: 301 YDEFVEKLVERVKLLKVGEDDPLDPDTDMGPLINEEQYEKIQSYYIEEAKAEGAKLVC G 360

Query: 372 D 372

Sbjct: 361 E 361

Fig. 5A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 8 of 43

8/43

Score = 219 (82.2 bits), Expect = 4.9e-15, P = 4.9e-15
Identities = 75/236 (31%), Positives = 105/236 (44%)

Query: 550 SSGNLHGYVAEGGAKDIRGAVEAAHQAFPG--WAGQSP-GXXXXXXXXXXXXERRKSTL 606
+ + G + V E + D+ AVEAA + AF G W SP E K L
Sbjct: 20 TNGEVIAQVPEATKEDVDKAVEAAREAFKGGEWGKTSPSERARILRKLADLLEENKDEL 79

Query: 607 AS--RLERQGXXXXXXXXXXXXXXRLRAW-GARVQAQGH-TLQVAGLRGP---VLRLRE 659
A+ LE LR + G + G T+ + P + RE
Sbjct: 80 AALETLETGKPLAEAKVAEVARAVDYLRYYAGMAEKLGEETIPTSLSESPGSMSYTMRE 139

Query: 660 PLGVLA VVCPDEWPLLAFVSLLAPALATGNTVV MVPSAACPLLAL---EVCQDMATVFPA 716
PLGV+A + P + PL+ V + APALA GNTVV+ PS PL AL E+ ++ P
Sbjct: 140 PLGVVAAITPWNFPLMMAVWKIAPALAA GNTVVLKPSEQTPLTALLAELIKEAEAGLPP 199

Query: 717 GLANVVTG-DRDHTRCLALHQDVQAMWYFGSAQ-GSQFVEWASAGNLKPVWASRG 770
G+ NVVTG + L D D+ + + GS + G ++ A+ NLKPV G
Sbjct: 200 GVINVVTGFGGAEVGEALVSHPDIDKISFTGSTEVGKAIMAAAENLKPVTLELG 255

Fig. 5B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 9 of 43
 9/43

Input file Fbh21481FL.seq; Output File 21481.trans
 Sequence length 1379

TTTGGCCCTCGAGGCCAAGAACATTGGCACGAGGAGCAAGTGGCCTAACACATGGATTTCTTCCAAAAATGCAGACCC
 ATTTTAATTAAGTTGTAAATTAACCCTGGGAGGGCAGGCCCTGGATCGGTCTGCTTCGGAGACACTGTGAGTA
 ACTTCCTATTGTTAACATTGGGATTAGCACGCCACTGGTGTTCAGCTGGAGGCTGCACAGAGCTGAGCTCC
 CTGCAGCCTTGGGCCTCCCCCTGCCCTGGAGTCCTGATCAGCGTCTTTGCAAAGCCAATCCCCTTTACTCCGTTG
 M G V M A M L M L P L L L L G I 16
 TCCCCCAGAACAAAG ATG GGA GTC ATG GCC ATG CTG ATG CTC CCC CTG CTG CTG CTG GGA ATC 48
 S G L L F I Y Q E V S R L W S K S A V Q 36
 AGC GGC CTC CTC TTC ATT TAC CAA GAG GTG TCC AGG CTG TGG TCA AAG TCA GCT GTG CAG 108
 N K V V V I T D A I S G L G K E C A R V 56
 AAC AAA GTG GTG GTG ATC ACC GAT GCC ATC TCA GGA CTG GGC AAG GAG TGT GCT CGG GTG 168
 F H T G G A R L V L C G K N W E R L E N 76
 TTC CAC ACA GGT GGG GCA AGG CTG GTG CTG TGT GGA AAG AAC TGG GAG AGG CTA GAG AAC 228
 L Y D A L I S V A D P S K T F T P K L V 96
 CTA TAT GAT GCC TTG ATC AGC GTG GCT GAC CCC AGC AAG ACA TTC ACC CCA AAG CTG GTC 288
 L L D L S D I S C V P D V A K E V L D C 116
 CTG TTG GAC CTC TCA GAC ATC AGC TGT GTC CCA GAT GTG GCA AAA GAA GTC CTG GAT TGC 348
 Y G C V D I L I N N A S V K V K G P A H 136
 TAT GGC TGT GTG GAC ATC CTC ATC AAC AAT GCC AGT GTG AAG GTG AAG GGG CCT GCC CAT 408
 K I S L E L D K K I M D A N Y F G P I T 156
 AAG ATT TCT CTG GAG CTC GAC AAA AAG ATC ATG GAT GCC AAT TAC TTT GGC CCC ATC ACA 468
 L T K A L L P N M I S R R T G Q I V L V 176
 TTG ACG AAA GCC CTG CTT CCC AAC ATG ATC TCC CGG AGA ACA GGC CAA ATC GTG TTA GTG 528
 N N I Q G K F G I P F R T T Y A A S K H 196
 AAT AAT ATC CAA GGG AAG TTT GGA ATC CCG TTC CGT ACG ACT TAC GCT GCC TCC AAG CAC 588

Fig. 6A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 10 of 43

10/43

A	A	L	G	F	F	D	C	L	R	A	E	V	E	E	Y	D	V	V	I	216
GCA	GCC	CTG	GGC	TTC	TTT	GAC	TGC	CTC	CGA	GCC	GAA	GTG	GAG	GAA	TAC	GAT	GTT	GTC	ATC	648
S	T	V	S	P	T	F	I	R	S	Y	H	V	Y	P	E	Q	G	N	W	236
AGC	ACC	GTG	AGC	CCG	ACT	TTC	ATC	CGG	TCG	TAC	CAC	GTG	TAT	CCA	GAG	CAA	GGA	AAC	TGG	708
E	A	S	I	W	K	F	F	F	R	K	L	T	Y	G	V	H	P	V	E	256
GAA	GCT	TCC	ATT	TGG	AAA	TTC	TTT	TTC	AGG	AAG	CTG	ACC	TAC	GGC	GTG	CAC	CCA	GTA	GAG	768
V	A	E	E	V	M	R	T	V	R	R	K	K	Q	E	V	F	M	A	N	276
GTG	GCG	GAG	GAG	GTG	ATG	CGC	ACC	GTG	CGG	AGG	AAG	AAG	CAA	GAG	GTG	TTT	ATG	GCC	AAC	828
P	I	P	K	A	A	V	Y	V	R	T	F	F	P	E	F	F	F	A	V	296
CCC	ATC	CCC	AAG	GCC	GCC	GTG	TAC	GTC	CGC	ACC	TTC	TTC	CCG	GAG	TTC	TTT	TTC	GCC	GTG	888
V	A	C	G	V	K	E	K	L	N	V	P	E	E	G	*					312
GTG	GCC	TGT	GGG	GTG	AAG	GAG	AAG	CTC	AAT	GTC	CCG	GAG	GAG	GGG	TAA					936
CTGCAGGAGGCCAAATGGGCCACCCCTTGGAAATAAAGGTTTCTGGCAAAAAAAAAAAAAAANTTGCAGG																				
CGCAAGCTTATTCCCTTAGGGAGGGTTAATTT																				

Fig. 6B

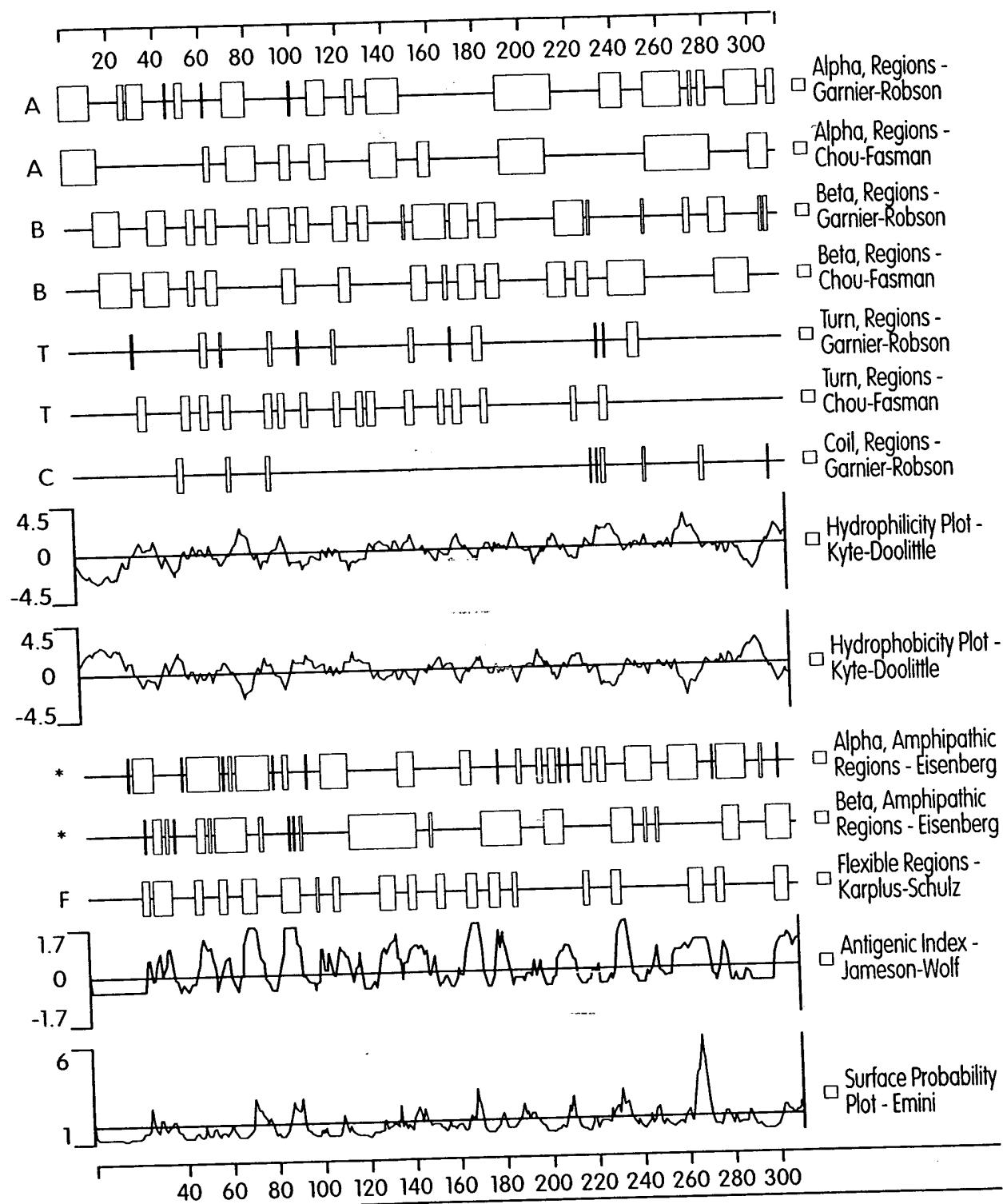


Fig. 7

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 12 of 43

12/43

Signal Peptide Predictions for 21481

Method	Predict	Score	Mat@
SignalP (eukaryote)	YES		19

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
7	23	ins-->out	4.5

>21481
MGVMAMLMPLLLGISGLLFIYQEVSRLWSKSAVQNKVVVITDAISGLGKECARVFHTG
GARLVLCGKNWERLENLYDALISVADPSKTFTPVLVLLDSLSDISCVPDVAKEVLD
CYGCV
DILINNASVKVKGPAHKISLELDKKIMDANYFGPITLTKALLPNMISRTGQIVLVNNIQ
GKFGIPFRTTYAASKHAALGFFDCLRAEVEEYDVVISTVSPTFIRSYHVP
EQGNWEASI
WKFFFRKLTYGVHPVEAEV
MRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFF
AVVACG
VKEKLNVPEEG

Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
265	283	ins-->out	0.2

>21481 mature
LLFIYQEVSRLWSKSAVQNKVVVITDAISGLGKECARVFHTGGARLVL
CGKNWERLENLY
DALISVADPSKTFTPVLVLLDSLSDISCVPDVAKEVLD
CYGCV
DILINNASVKVKGPAHKISLELDKKIMDANYFGPITLT
KALLPNMISRTGQIVLVNNIQ
GKFGIPFRTTYAASKHAALGFFDCLRAEVEEYDV
VISTVSPTFIRSYHVP
EQGNWEASI
WKFFFRKLTYGVHP
VEAEV
MRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFF
AVVACG
VKEKLNVPEEG

Fig. 8

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 13 of 43
 13/43

Protein Family / Domain Matches, HMMer Version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.0/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.9650.seq

Query: 21481

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
adh_short	short chain dehydrogenase	120.0	4.5e-32	1
A2M	Alpha-2-macroglobulin family	0.5	7.1	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
adh_short	1/1	38	227	..	1	203	[] - 120.0 4.5e-32
A2M	1/1	278	291	..	1	14	[.] 0.5 7.1

Alignments of top-scoring domains:

adh_short: domain 1 of 1, from 38 to 227: score 120.0, E = 4.5e-32

*->KvaLvTGassGIGlaiAkrLakeGakVvvadrneeklekGavakelk
 Kv++T a sG+G++A+ +++ Ga++v+++ n e+le+ ++

21481 38 KVVVITDAISGLKECARVFHTGGARLVLCGKNWERLEN--LYDALI 82

e1Ggnd..kdrAlaiql1Dvtdeesv.aaveqaverlGr1DvLVNNAGgii
 +++++ + 1D++d + v++++++ +G +D+L+NMA +

21481 83 SV-ADPsktFTPKLVLLDSLSDISCVpDVAKEVLDCYGCVDILINNAS--V 129

1lrpgpfaelsrtmeedwdrvividvNltgvflltravlplmamkkrggGrI
 gp+++++s +e+ ++++d N++g++ lt+a+lp m+ r+ G I

21481 130 -KVKGPAHKIS---LELDKKIMDANYFGPITLTКАLLP--NMISRRTGQI 173

vNiSSvaGrkegg1vgvpagsaYsASKaAvigltrsLAElaphgIrVna
 v + + G + g p+++ Y+ASK+A g+ ++L+ E+ ++ + ++

21481 174 VLVNNIQG-----KFGIPFRTTYAASKHAALGFFDCLRAEVEYDVVIST 218

VAPGgvdTd<-*

v+P +++

21481 219 VSPTFIRSY 227

A2M: domain 1 of 1, from 278 to 291: score 0.5, E = 7.1

->idedditiRSyFPE<-

i+ + +R++FPE

21481 278 IPKAAVYVRTFFFPE 291

Fig. 9

14/43

ProDom Matches

ProdomId	Start	End	Description	Score
View Prodom 11 <input type="button" value="Boxer ▾"/> <input type="button" value="Showing match ▾"/> <input type="button" value="Go!"/>	99	219	p99.2 (1078) ADH(34) GALE(20) FABG(13) // OXIDOREDUCTASE PROTEIN DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIS SYNTHASE ALCOHOL PUTATIVE	113
ProdomId	Start	End	Description	Score

View Prodom 11

>11 p99.2 (1078) ADH(34) GALE(20) FABG(13) // OXIDOREDUCTASE PROTEIN
DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIS SYNTHASE ALCOHOL PUTATIVE
Length = 269

Score = 113 (44.8 bits), Expect = 0.00016, P = 0.00016
Identities = 41/138 (29%), Positives = 63/138 (45%)

Query: 99 DLSDIS-CVPDVAKEVLD CYGCVDILINNASVKV-KGPAHKISLELD-----KKIMDANY 151
 D+ D+ V V +E +G +D+L+NNA V K A ++ E +++++ N
 Sbjct: 87 DVEDVEKLVETVVEEFSGIHGKIDVLVNNAGVMAPKAVAESMTEETSDEEWEVIEVN 146

Query: 152 FGPIITLT KALLPNMIS----- RRTGQIVLVNNIQGK-FGIP-FRTTYAASKHAALGF 201
 G LT+A LP M R G IV V ++ G G P + Y+ASK A F
 Sbjct: 147 TGTFLNTQAALPAMKKFSDAAKKRFVGTVTIVNVASVAGSTMGSQAYSASKAAVESF 206

Query: 202 FDCLRAEVEEYDVVISTV 219
 L E+ Y ++ V
 Sbjct: 207 TKSLAMELSPYSASVAMV 224

Fig. 10

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 15 of 43
 15/43

Input file Fbh25964F1.seq; Output File 25964.trans
 Sequence length 1725

GAGAAGGAGGAGCCAGCGGAAGGACGGTGTGCGGGCCGGCAGCCCTGGACGAAAGAAGAGGGCCCTCCAGGCCAGTC
 TGGGCACCTGGGATAGCGGCTGCAGCCATCAGCAGGGCAGACGGCAGGTGGCTGGTTGCTGCAGCTCCAGGATCA
 GCTCTGCCCTCCCCGCAAACGCCAGCCTCGTACCGCTCCAGGCACCTCCAGCAGTAACAGGTGGTGCAGCAGGTGG

 M A D S A Q A Q K 9
 CAGCCAGCCCCTGGATGAGCCAAGGTCTCTTCCCCAGCCAGGC ATG GCC GAC TCT GCA CAG GCC CAG AAG 27

 L V Y L V T G G C G F L G E H V V R M L 29
 CTG GTG TAC CTG GTC ACA GGG GGC TGT GGC TTC CTG GGA GAG CAC GTG GTG CGA ATG CTG 87

 L Q R E P R L G E L R V F D Q H L G P W 49
 CTG CAG CGG GAG CCC CGG CTC GGG GAG CTG CGG GTC TTT GAC CAA CAC CTG GGT CCC TGG 147

 L E E L K T G P V R V T A I Q G D V T Q 69
 CTG GAG GAG CTG AAG ACA GGG CCT GTG AGG GTG ACT GCC ATC CAG GGG GAC GTG ACC CAG 207

 A H E V A A A V A G A H V V I H T A G L 89
 GCC CAT GAG GTG GCA GCA GCT GTG GCC GGA CAT GTG GTC ATC CAC ACG GCT GGG CTG 267

 V D V F G R A S P K T I H E V N V Q G T 109
 GTA GAC GTG TTT CCC AGG GCC AGT CCC AAG ACC ATC CAT GAG GTC AAC GTG CAG GGT ACC 327

 R N V I E A C V Q T G T R F L V Y T S S 129
 CGG AAC GTG ATC GAG GCT TGT GTG CAG ACC GGA ACA CGG TTC CTG GTC TAC ACC AGC AGC 387

 M E V V G P N T K G H P F Y R G N E D T 149
 ATG GAA GTT GTG GGG CCT AAC ACC AAA GGT CAC CCC TTC TAC AGG GGC AAC GAA GAC ACC 447

 P Y E A V H R H P Y P C S K A L A E W L 169
 CCA TAC GAA GCA GTG CAC AGG CAC CCC TAT CCT TGC AGC AAG GCC CTG GAG TGG CTG 507

 V L E A N G R K V R G G L P L V T C A L 189
 GTC CTG GAG GCC AAC GGG AGG AAG GTC CGT GGG GGG CTG CCC CTG GTG ACG TGT GCC CTT 567

 R P T G I Y G E G H Q I M R D F Y R Q G 209
 CGT CCC ACG GGC ATC TAC GGT GAA GGC CAC CAG ATC ATG AGG GAC TTC TAC CGC CAG GGC 627

 L R L G G W L F R A I P A S V E H G R V 229
 CTG CGC CTG GGA GGT TGG CTC TTC CGG GCC ATC CCG GCC TCT GTG GAG CAT GGC CGG GTC 687

Fig. 11A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 16 of 43

16/43

Y	V	G	N	V	A	W	M	H	V	L	A	A	R	E	L	E	Q	R	A	249
TAT	G TG	GGC	A AT	G TT	GCC	T GG	A TG	C AC	G TG	C TG	G CA	G CC	C GG	G AG	C TG	G AG	C AG	C GG	G CA	747
A	L	M	G	G	Q	V	Y	F	C	Y	D	G	S	P	Y	R	S	Y	E	269
GCC	CTG	ATG	GGC	GGC	CAG	GTA	TAC	TTC	TGC	TAC	GAT	GGA	TCA	CCC	TAC	AGG	AGC	TAC	GAG	807
D	F	N	M	E	F	L	G	P	C	G	L	R	L	V	G	A	R	P	L	289
GAT	TTC	AAC	ATG	GAG	TTC	CTG	GGC	CCC	TGC	GGA	CTG	CGG	CTG	GTG	GGC	GCC	CGC	CCA	TTG	867
L	P	Y	W	L	L	V	F	L	A	A	L	N	A	L	L	Q	W	L	L	309
CTG	CCC	TAC	TGG	CTG	CTG	GTG	TTC	CTG	GCT	GCC	CTC	AAT	GCC	CTG	CTG	CAG	TGG	CTG	CTG	927
R	P	L	V	L	Y	A	P	L	L	N	P	Y	T	L	A	V	A	N	T	329
CGG	CCA	CTG	GTG	CTC	TAC	GCA	CCC	CTG	CTG	AAC	CCC	TAC	ACG	CTG	GCC	GTG	GCC	AAC	ACC	987
T	F	T	V	S	T	D	K	A	Q	R	H	F	G	Y	E	P	L	F	S	349
ACC	TTC	ACC	GTC	AGC	ACC	GAC	AAG	GCT	CAG	CGC	CAT	TTC	GGC	TAT	GAG	CCC	CTG	TTC	TCG	1047
W	E	D	S	R	T	R	T	I	L	W	V	Q	A	A	T	G	S	A	Q	369
TGG	GAG	GAT	AGC	CGG	ACC	CGC	ACC	ATT	CTC	TGG	GTA	CAG	GCC	GCT	ACG	GGT	TCA	GCC	CAG	1107
*																				370
TGA																				1110

CGGTGGGCTGGGCCCTGGAGGCCAGATAACAGCACATCCACCCAGGTCCCGAGCCCTCACACCCCTGGACGGGAAGGGA
 CAGCTGCATTCCAGAGCAGGAGGCAGGGCTGGGCCAGAATGGCTGCTTGCTAGAGCCCTCCACATTTCTTT
 TTCTTTTTGAGACAGGGCTTGCTCTGTCACCCAGACTGGAATGCAAGTGGTGTGANTCATAAGCTCACTNGMACCCT
 YAANCCTCTGGTTCAAGCAATCCTNCTNGCCTYAANCCTCTNGAACAGCTGGGANCCACAGGTGCACGCCANC
 CACANCCTGGCTTTTTTT

Fig. 11B

17/43

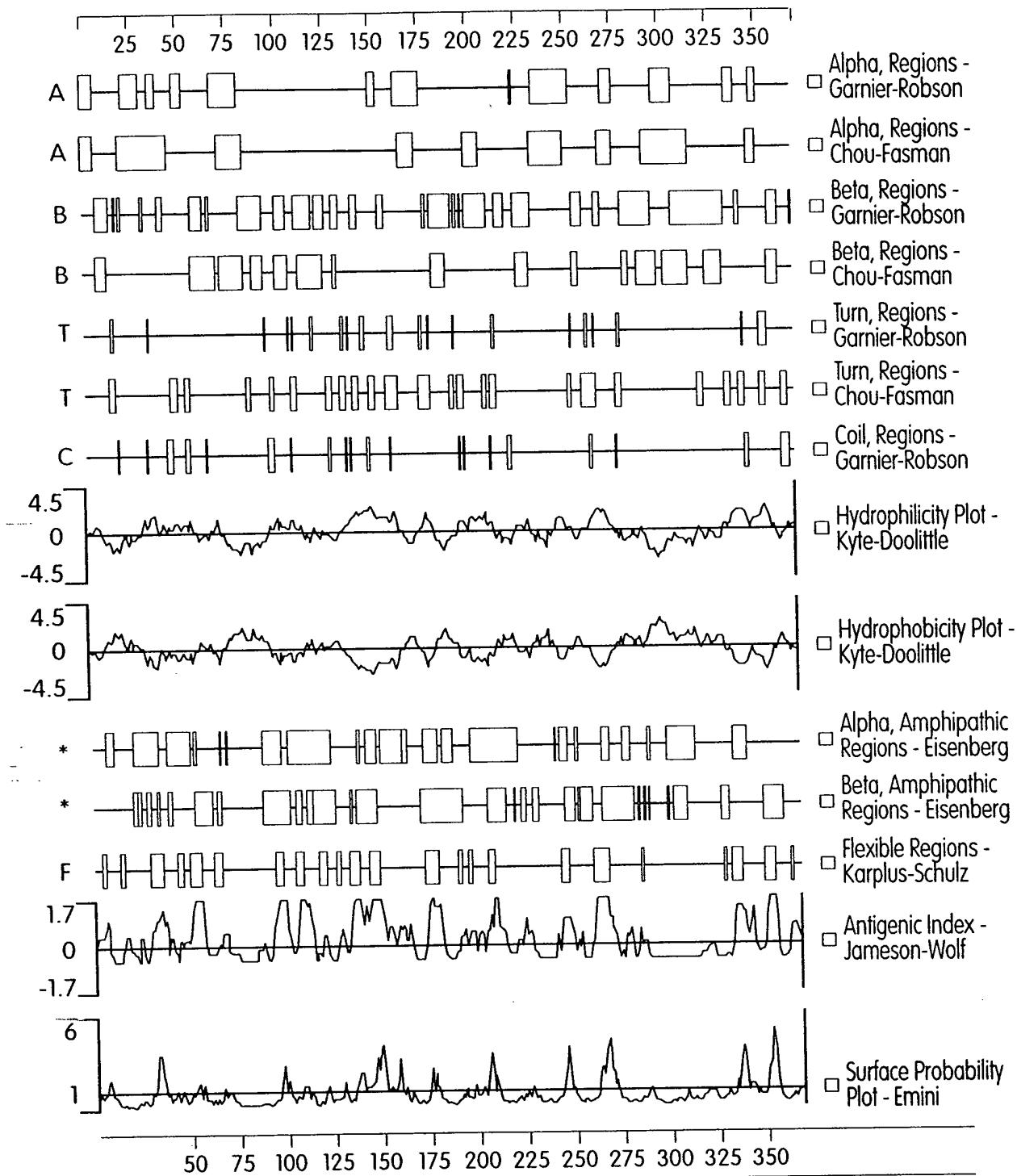


Fig. 12

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
10	26	out-->ins	1.0
73	90	ins-->out	2.0
289	305	out-->ins	3.0
312	333	ins-->out	1.2

>25964
MADSAQAQKLVYLVTGGCGFLGEHVVRMLLQREPRLGELRVFDQHLGPWLEELKTGPV
RVTAIQGDVTQAHEVAAAVAGAHVVIHTAGLVDVGRASP
KTIHEVNQGTRNVIEACVQTGTRFLVYTSSMEVVGPN
TKGHPFYRGNEDTPYEAVHRHPYPCSKALAELVLEANGRK
VRGGLPLVTCALRPTGIYGEHQIMRDFYRQLRLGGWL
FRAIPASVEHGRVYVGNVAWMHVL
AARELEQRAALMGGQVYFCYDGSPYRSYEDFN
MEFLGPCGLRLVGARPLLPYWLLVFLAA
LNALLQWLLRPLVLYAPLLNPYTLAVANTTFTVSTD
KAQRHF
GYEPLFSWEDSRRTILW
VQAATGSAQ

Fig. 13

Applicants: Rachel E. Meyers, et al.
 Title: 21481. A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
 Attorney/Agent: Kerri Pollard Schray
 Docket No.: MPI00-079P1RCP2CN1M
 Sheet 19 of 43
 19/43

Protein Family / Domain Matches, HMMer Version 2

Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.0/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.9289.seq

Query: 25964

Scores for sequence family classification (score includes all domains):				
Model	Description		Score	E-value
3Beta_HSD	3-beta hydroxysteroid dehydrogenase/iso	676.9	1e-199	1
S-AdoMet_synt	S-adenosylmethionine synthetase	1.8	0.78	1
adh_short	short chain dehydrogenase	-48.6	0.022	1
Epimerase	NAD dependent epimerase/dehydratase fam	-148.0	0.0016	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
adh_short	1/1	10	197	..	1 203 []	-48.6	0.022
S-AdoMet_synt	1/1	341	351	..	365 376 .]	1.8	0.78
3Beta_HSD	1/1	1	365	[.	1 425 []	676.9	1e-199
Epimerase	1/1	12	365	..	1 359 []	-148.0	0.0016

Alignments of top-scoring domains:

adh_short: domain 1 of 1, from 10 to 197: score -48.6, E = 0.022
 *->KvaLviGassGIGlaiAkrLakeGakVvvadrneeklekGavalelk
 v LvTG+++ +G +++ L+ + ++ ++ + G +++elk
 25964 10 LVYLVTGGCGFLGEHVVVRMLLQR--EPRLGELRVFDQHLGPWLEELK 54

elGgndkdralaiqlDvtdeesv.aaveqaverlGr1DvLVNNAGgiill
 + r+ aiq+Dvt++ +v aav+ a +v++ AG +
 25964 55 TGPV---RVTAIQGDVTQAHEVaAAVAGA-----HVVIAHTAG--L-- 89

rpgfaelsrtmeedwdrvividvNltgvflitravlplmamkkrggGrIvN
 + f + s ++ +++vN+ g tr v++ a ++ g v
 25964 90 -VDVFGRAS---PK---TIHEVNQG---TRNVIE--ACVQTGTRFLVY 126

iSSvaGrke.....g.glvvgpggsaYsASKaAvigltrs
 +SS +e +++++++ +++ + + ++ +Y +SKa l++
 25964 127 TSS---MEvvgpntkghpfyrgnEdTPYEAVHRHPYPCSKA---LAEW 168

LAlElaphgIr.....VnavapGgvdTd<-*
 L 1E +++r++ + a P g++ +
 25964 169 LVLEANGRKVRggplvTCALRPTGIYGE 197

S-AdoMet_synt: domain 1 of 1, from 341 to 351: score 1.8, E = 0.78
 ->HFGreevdFpWE<-

HFG e F+WE
 25964 341 HFGYEP-LFSWE 351

Fig. 14A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 20 of 43
 20/43

3Beta_HSD: domain 1 of 1, from 1 to 365: score 676.9, E = 1e-199
 *->elsesldmaglsclVTGGgGF1GrhIVreLlregeslqevRvfDlrf
 +++s++ l++lVTGG+GF1G+h Vr+Ll+++++l e+RvfD +
 25964 1 -MADSAQAKLVYLVTGGCGFLGEHVVRLLQREPRLGELRVFDQHL 46

 spelde.dssklqvitkikyieGDvtDkqdlaaAlqgiSCCTLDMTLmD
 +p+l+e +++++ v+ +i+GDvt+++++aaA++g+
 25964 47 GPWLEELKTGPVRVT---AIQGDVTQAHEVAAAAGA----- 80

 dvvIHtAaiiDvfGelrvsGSDLGSFGVTVLFLAVTEGSYVVFYmGATDLR
 +vvIHtA+++DvfG
 25964 81 HVIHTAGLVDVFG----- 94

 kasrdrimkNVkGTqnvldACveaGVrvlVYTSSmeVVGpNsrgqpivN
 as+ i++VNV+GT+nv++ACv++G+r+lVYTSSmeVVGpN +G+p+++
 25964 95 RASPPTIHEVNQGTRNVIEACVQTGTRFLVYTSSMEVVGPNTKGHPFYR 144

 GdEttPYestDDhqdaYpeSKalAEk1VLkANGsm1knGgrLyTCALRPa
 G+E+tpYe++ h+++Yp+SKalAE 1VL+ANG+ +++G L+TCALRP+
 25964 145 GNEDTPYEAV--HRHPYPCSKALAEWLVLLEANGRKVRGGPLVTCALRPT 192

 gIfGeGdqflvpflrqlvknGlakfriGdknalsdrvYVgNVAwAHILAA
 gI+GeG q + f+rq ++G+ fr ++ rVYVgNCaw+H+LAA
 25964 193 GIYGEGHQIMRDFYRQGLRLGGWLFRAlPASVEHGRVYVGNAWMHVLA 242

 raLqdppkkGREGassiaGqaYFIsDdsPvnSYddFnrtllkalGlrlpst
 r+L+++ a+ Gq+YF++D+sP++SY+dFn+++l ++Glrl +
 25964 243 RELEQR-----AALMGGQVYFCYDGSPYRSYEDFNMEFLGPCGLRLVGA 286

 w.r1PlpllyvlaynellswLrk1alrYtP1lnpytvtnattFtfst
 ++lP++ll++la+ln+ll+wLLr+l + Y P1lnpyt+++anttFt+st
 25964 287 RpLLPYWLLVFLAALNALLQWLLRPL-VLYAPLLNPYTLAVANTTFTVST 335

 nKAkkdLGYePlvtwEEarakTieWiqeLe<-*
 +KA+++GYePl++WE +r +Ti+W+q+
 25964 336 DKAQRHFGYEPPLFSWEDSRTRTILWVQAAT 365

Fig. 14B-1

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 21 of 43

21/43

Epimerase: domain 1 of 1, from 12 to 365: score -148.0, E = 0.0016

*->ILVTGGAGFIGShlvreLnn...ygddkVvvLDnLtdyYqyagnea
+LVTGG GF G h+vr Li+ +++ g +v + + +
25964 12 YLVTGGCGFLGEHVVRLQReprLGELRV----FD----QHLGPW 49

rlevvegnprytFvkGDIcDrlddkvfaehqpDaViHfAAeshV.drSi
++e + g r+t ++GD+ + + ++a +ViH A++ V +r
25964 50 LEEELKTGPVRVTAIQGDVTQAHEVAAAAGA--HVVVIHTAGLVDFGR-- 95

ekPlayidtNvvGTltLLEaaRnyWsaLdetkagvkkfvfsSTdeVYGdI
P + + Nv GT + +Ea+ g +v+ S+ eV G +
25964 96 ASPKTIHEVNQGTRNVIEACV-----QTGTRFLVYTSSMEVVGPN 136

esiPisaF...tEdtPynPs..SPYgaSKassEllvrayhgraygLpailL
++ + F ++ EdtPy ++ PY SKa E lv +
25964 137 TKGHP--FyrgNEDTPYEAVhrHPYPCSKALAEWLLEAN----- 174

RyFNvYGPYqsgriGedpngfpekLIPliliqnalgkgeplpvYGdDYpTp
G+ g+ +Pl+ + al p +YG
25964 175 -----GRKVRRGG-----LPLV-TCALR---PTGIYG----- 196

DGtqv.RDw.....ihVeDharANhillaltkg
+G q+ RD+ +++ + ++ + + ++++++V ++a h+la +++
25964 197 EGHQImRDFyrqglrlggwlfraipasvehgrVYVGNVAWM-HVLAAREL 245

.....raGkgsevYNiGg
+++ +++ +++++ ++ ++ + + ++ + + +G
25964 246 eqraalmggqvyfcydgspyrsyedfnmeflgpcqlrLVG----- 285

gneysnlEvVeaIekllgelaPekphvkakedpatfvddRpGddarya..
+ + + + ++++++l+ 1 + ++ +++ ++++++a
25964 286 -ARPLLPYLLVFLAALNALLQWL-----LRPLVLYAPLLN--PYTLAva 327

....aDasKikreLGWkPevtnleeGladTvnWylene<-*
+++ +++ +K++I G++P + e+ +T+ W +
25964 328 nttftVSTDKAQRHFGYEPLFS-WEDSRTRTILWVQAAAT 365

Fig 14B-2

22/43

ProdomId	Start	End	Description	Score
View Prodom 1280 <input type="button" value="Boxer ▼"/>			p99.2 (39) 3BHS(5) 3BH1(4) 3BH2(3) // DEHYDROGENASE STEROID BETA-HYDROXYSTEROID 3BETA-HSD DEHYDROGENASE/DELTA 5-->4-ISOMERASE INCLUDES: PROGESTERONE 3-BETA-HYDROXY-DELTA5-STEROID 3-BETA-HYDROXY-5-ENE	395
Showing match <input type="button" value="▼"/> <input type="button" value="Go!"/>	11	362		

View Prodom 1280

>1280 p99.2 (39) 3BHS(5) 3BH1(4) 3BH2(3) // DEHYDROGENASE STEROID BETA-HYDROXYSTEROID 3BETA-HSD DEHYDROGENASE/DELTA 5-->4-ISOMERASE INCLUDES: PROGESTERONE 3-BETA-HYDROXY-DELTA5-STEROID 3-BETA-HYDROXY-5-ENE
Length = 416

Score = 395 (144.1 bits), Expect = 3.2e-42, Sum P(2) = 3.2e-42
Identities = 99/268 (36%), Positives = 134/268 (50%)

Query: 102 HEVNVQGTRNVIEACVQTGTRFLVYTSSMEVVGPNTKGHPFYRGNEDTPYEAVHRHPYPC 161
 ++ NVQGTRN+IE C RF MEV GPN+ G+E+ +E+ +PYP
 Sbjct: 157 YKFNVQGTRNLIEKC-----RFF---GVMEVAGPNSYKEIILNGHEEEHHESTWPNPYPY 208

 Query: 162 -SKALAEWLVLLEANGRKVRRGLPLVTCALRPTGIYGEHQIMRDFYRQGLRLGGWLFR 220
 SK +AE VL ANG ++ G L TCALRP IYGEQ + + Q L+ GG +FR
 Sbjct: 209 YSKKMAEKAVLAANGSMLKNNGTLYTCALRPMIYGEQDKFLSPMIVQALKNGGIMFRVG 268

 Query: 221 PASVEHGRVYVGNVAWMHVXXXXXXXXXX--MGGQVYFCYDGSPYRSYEDFNMEFLGP 278
 VYVGNVAW H+ + GQ Y+ D +P++SY+D N
 Sbjct: 269 GKFSVANPVYVGNVAWAHILAARGLQDPKKSPNIQQFYYISDDTPHQSYDDLNYTLSKE 328

 Query: 279 CGLRLVGARPLP----YWXXXXXXXXXXXXXXXXXXXXPYTLAVANTTFVS 334
 GLRL ++ LP YW N + + ++NTTFT S
 Sbjct: 329 WGLRLDSSKWRLPLLLYWLAFLLEMVSFLRPISYNQPPF---NRHLVTLSNTTFTFS 385

 Query: 335 TDKAQRHFGYEPLFSWEDSRRTILWVQ 362
 KAQR GYEPL SWE+++ +T W++
 Sbjct: 386 YKKAQRDLGYEPLVSSEEAKQKTSEWIE 413

Score = 65 (27.9 bits), Expect = 3.2e-42, Sum P(2) = 3.2e-42
Identities = 11/23 (47%), Positives = 17/23 (73%)

Query: 11 VYLVTTGGCGFLGEHVRMLLQRE 33
 VY VTGG FLG ++V++L+ +
 Sbjct: 14 VYAVTGGAEFLGRYIVKLISAD 36

Fig. 15

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 23 of 43

23/43

Input file Fbh21686F1.seq; Output File 21686.trans
 Sequence length 1209

	M	S	L	R	4
CCCACGCGTCCGCCACCGTCCGGACCGTGGCGGACCGTGGCGCCGCCTGA	ATG	TCC	CTG	AGA	12
P R R A C A Q L L W H P A A G M A S W A					24
CCC AGA AGG GCC TGC GCT CAG CTG CTC TGG CAC CCC GCT GCA GGG ATG GCC TCC TGG GCT					72
K G R S Y L A P G L L Q G Q V A I V T G					44
AAG GGC AGG AGC TAC CTG GCG CCT GGT TTG CTG CAG GGC CAA GTG GCC ATC GTC ACC GGC					132
G A T G I G K A I V K E L L E L G S N V					64
GGG GCC ACG GGC ATC GGA AAA GCC ATC GTG AAG GAG CTC CTG GAG CTG GGG AGT AAT GTG					192
V I A S R K L E R L K S A A D E L Q A N					84
GTC ATT GCA TCC CGT AAG TTG GAG AGA TTG AAG TCT GCG GCA GAT GAA CTG CAG GCC AAC					252
L P P T K Q A R V I P I Q C N I R N E E					104
CTA CCT CCC ACA AAG CAG GCA CGA GTC ATT CCC ATA CAA TGC AAC ATC CGG AAT GAG GAG					312
E V N N L V K S T L D T F G K I N F L V					124
GAG GTG AAT AAT TTG GTC AAA TCT ACC TTA GAT ACT TTT GGT AAG ATC AAT TTC TTG GTG					372
N N G G Q F L S P A E H I S S K G W H					144
AAC AAT GGA GGA GGC CAG TTT CTT TCC CCT GCT GAA CAC ATC AGT TCT AAG GGA TGG CAC					432
A V L E T N L T G T F Y M C K A V Y S S					164
GCT GTG CTT GAG ACC AAC CTG ACG GGT ACC TTC TAC ATG TGC AAA GCA GTT TAC AGC TCC					492
W M K E H G G S I V N I I V P T K A G F					184
TGG ATG AAA GAG CAT GGA GGA TCT ATC GTC AAT ATC ATT GTC CCT ACT AAA GCT GGA TTT					552
P L A V H S G A A R A G V Y N L T K S L					204
CCA TTA GCT GTG CAT TCT GGA GCT GCA AGA GCA GGT GTT TAC AAC CTC ACC AAA TCT TTA					612
A L E W A C S G I R I N C V A P G V I Y					224
GCT TTG GAA TGG GCC TGC AGT GGA ATA CGG ATC AAT TGT GTT GCC CCT GGA GTT ATT TAT					672
S Q T A V E N Y G S W G Q S F F E G S F					244
TCC CAG ACT GCT GTG GAG AAC TAT GGT TCC TGG GGA CAA AGC TTC TTT GAA GGG TCT TTT					732
Q K I P A K R I G V P E E V S S V V C F					264
CAG AAA ATC CCC GCT AAA CGA ATT GGT GTT CCT GAG GAG GTC TCC TCT GTG GTC TGC TTC					792
L L S P A A S F I T G Q S V D V D G G R					284
CTA CTG TCT CCT GCA GCT TCC TTC ATC ACT GGA CAG TCG GTG GAT GTG GAT GGG GGC CGG					852
S L Y T H S Y E V P D H D N W P K G A G					304
AGT CTC TAT ACT CAC TCG TAT GAG GTA CCA GAT CAT GAC AAC TGG CCC AAG GGA GCA GGG					912
D L S V V K K M K E T L K E K A K L *					323
GAC CTT TCT GTT GTC AAA AAG ATG AAG GAG ACC TTA AAG GAG AAA GCT AAG CTC TGA					969
GCTGAGGAAACAAGGTGTCCATCCCCAGTGCCTCACATCTGAGGATATGCTTCTGTACTTTTAAAGCTTATA					
GTTGGTATGGAAAACATTTCCTATTTTAAGTGTATTAAATTATCTATGGAAAAACTATTCCCTGAAATATATACA					
GTCTTATGTCCCCAAAAAA					

Fig. 16

CLUSTAL W (1.74) multiple sequence alignment

5052204_SDR_rat 21686	-----MGSWKSGQSYLAAGLLQNQVAVVTGGATGIGKAISRELLHL MSLRPRRACAQLLWHPAAGMASWAKGRSYLAPGLLQGQVAIVTGGATGIGKAIVKELLEL *.*** . *;*****.****.***:***** :***.*
5052204_SDR_rat 21686	GCNVVIASRKLDRLTAAVDELRASQPPSSSTQVTAIQCNIRKEEVNNLVKSTLAKYGKI GSNVVIASRKLERLKSADELQANLPPTKQARVPIQCNIRNEEVNNLVKSTLDTFGKI *.*****:***.:*.***:*. ***:.* .*****:***** . :***
5052204_SDR_rat 21686	NFLVNNAGGQFMAPAEDITAKGWQAVIETNLTTGTFTYMCKAVYNSWMKDHGGSIVNIIVLL NFLVNNNGGQFLSPAEHISSKGWHAVLETNLTTGTFTYMCKAVYSSWMKEHGGSIVNIIVPT ***** ,****:***,*;***:***:*****:***** .****:*****
5052204_SDR_rat 21686	NNGFPTAAHSGAARAGVYNLTKTMALTWASSGVRINCVAPGTIYSQTAVDNYGELGQTMF KAGFPLAVHSGAARAGVYNLTKSALLEWACSGIRINCVAPGVIYSQTAVENYGSWGQSFF : *** * .*****:*****:***:*** .***:***** .*****:***. ***:*
5052204_SDR_rat 21686	EMAFENIPAKRVGLPEEISPLVCFLSPAASFITCQLINV DGGQALYTRNFTIPDHNDWP EGSFQKIPAKRIGVPEEVSSVCFLLSPAASFITGQSVDVGGRSLYTHSYEVPDHDNWP * :*:*****:***:***:***:*****:*****:*****:***:.. :*****
5052204_SDR_rat 21686	VGAGDSSFIKKVKESLKKQARL KGAGDLSVVKKMKGTLKEKAKL **** * .***:***:***:***:*

Fig. 17

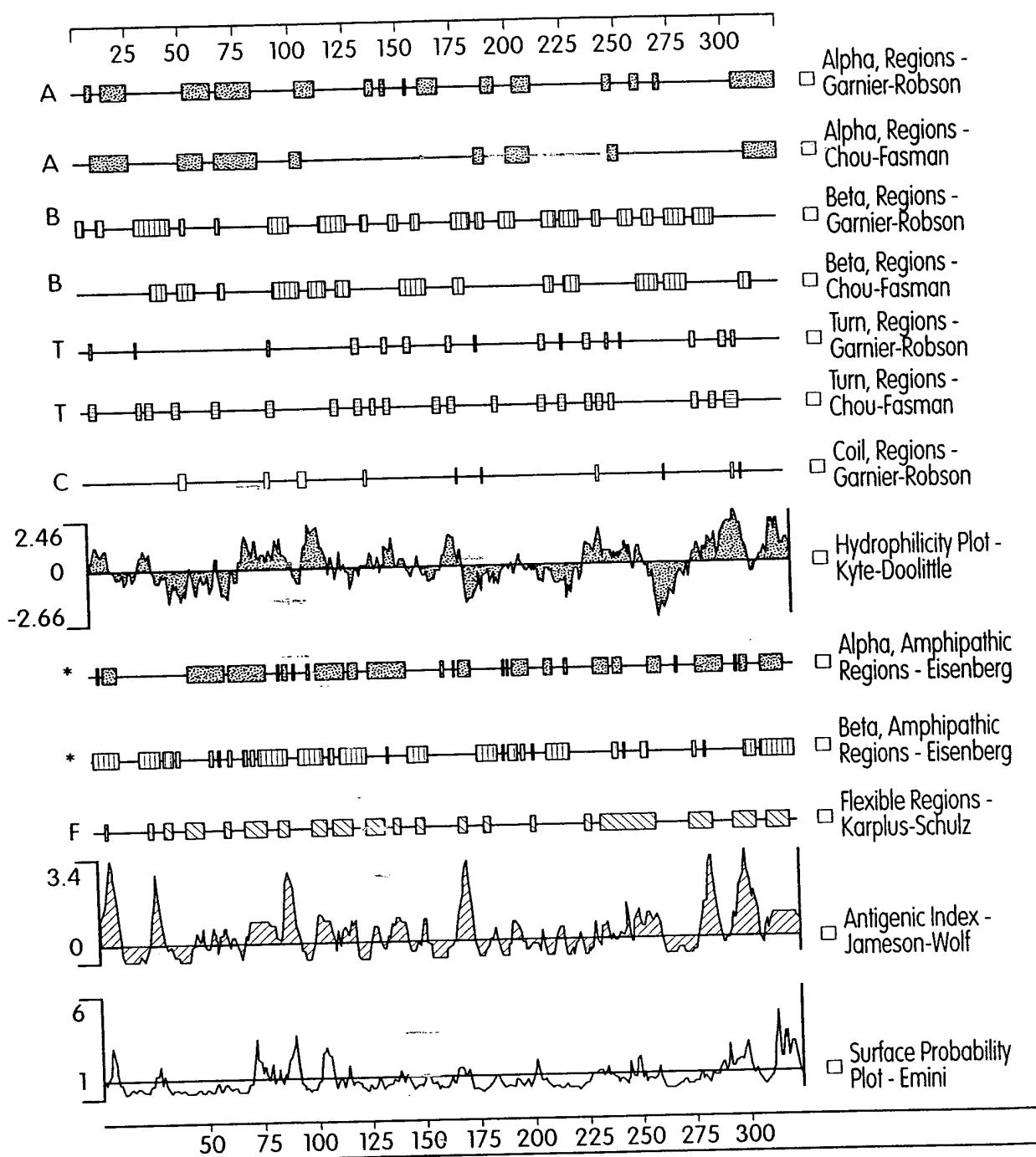


Fig. 18

26/43

Signal Peptide Predictions for 21686

Method	Predict	Score	Mat@
Signal (eukaryote)	MAYBE		20

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
29	50	ins-->out	0.9
170	188	out-->ins	0.2
208	224	ins-->out	0.6
258	275	out-->ins	2.6

>21686

MSLRPRRACAQLLWHPAAGMASWAKGRAYLAPGLLQGQVAIVTGGATGIGKAIVKELLEL
 GSNVVIASRKLERLKSAADELQANLPPTKQARVIPIQCNIRNEEVNNLVKSTLDTFGKI
 NFLVNNGGGQFLSPAEHISSKGWHAVLETNLTTGYMCKAVYSSWMKEHGGSIVNIIIVPT
 KAGFPLAVHSGAARAGVYNLTLSALEWACSGIRINCVAvgVIYSQTAVENYGSWGQSFF
 EGSFQKIPAKRIGVPEEVSSVVCFLSPAASFITGQSVVDGGRSLYTHSYEVPDHDNWP
 KGAGDLSVVKKMKETLKEKAKL

Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
10	31	ins-->out	0.9
151	169	out-->ins	0.2
189	205	ins-->out	0.6
239	256	out-->ins	2.6

>21686_mature

MASWAKGRSYLAPGLLQGQVAIVTGGATGIGKAIVKELLELGSNVVIASRKLERLKSAAD
 ELQANLPPTKQARVIPIQCNIRNEEVNNLVKSTLDTFGKINFLVNNGGGQFLSPAEHIS
 SKGWHAVLETNLTTGYMCKAVYSSWMKEHGGSIVNIIIVPTKAGFPLAVHSGAARAGNYN
 LTKSLALEWACSGIRINCVAvgVIYSQTAVENYGSWGQSFFEGSFQKIPAKRIGVPEEV
 SVVCFLSPAASFITGQSVVDGGRSLYTHSYEVPDHDNWPKGAGDLSVVKKMKETLKE
 AKL

Fig. 19

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 27 of 43

27/43

Protein Family / Domain Matches, HMMer Version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

 HMM file: /prod/ddm/seqanal/PFAM/pfam4.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.19160.seq

 Query: 21686

Scores for sequence family classification (score includes all domains):					
Model	Description		Score	E-value	N
<u>adh short</u>	short chain dehydrogenase		162.5	7.3e-45	1
<u>adh short_C2</u>	short chain dehydrogenase/reductase C-te		47.2	3.7e-10	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
<u>adh short</u>	1/1	38	226 ..	1	203 []	162.5	7.3e-45
<u>adh_short_C2</u>	1/1	250	280 ..	1	31 []	47.2	3.7e-10

Alignments of top-scoring domains:
adh_short: domain 1 of 1, from 38 to 226: score 162.5, E = 7.3e-45
 *->KvaLvTGassGIGlaiAkrLakeGakVvvadrneeklekGavakelk
 +va+vTG++ GIG+ai+k+l++ G +Vv+a r e+l +++
 21686 38 QVAIVTGGATGIGKAIVKELLELGSNVVIASRKLERL----KSAAD 79

elGgnd....kdralaiqlDvtdeesv.aaveqaverlGrlDvLVNNAGg
 el +n+++++ r+++iq++++ ee+v++v+ ++ +G+++ LVNN Gg
 21686 80 ELQANLpptkQARVIPIQCNIRNEEVnNLVKSTLBtFGKINFLVNNGGG 129

.iillrpgrfaelsrtmeedwdrvividvNltgvfltravlplmamkkrgg
 +++ p++ +s + w +v+++Nltg+f++++av +k +g
 21686 130 qFL---SPAEHIS---SKGWHAVLETNLTFYMCNAVYS--SWMKEHG 170

GrIvNiSSVvaGrkegg1vgvpggsaYsASKaAvigltrsLA1ElaphgIr
 G+IvNi + g+p ++ +A+ a+v lt+sLA1E+a gIr
 21686 171 GSIVNIIIV-PT----KAGFPLAVHSGAARAGVYNLTksLAleWACSGIR 214

VnavapGgvdTd<-*
 +n+vaPG ++ +
 21686 215 INCVAPGVIYSQ 226

adh_short_C2: domain 1 of 1, from 250 to 280: score 47.2, E = 3.7e-10
 ->gRlGePeEiAnavvFLASdaAsYiTGqt1vV<-
 +R G PeE++++v FL S+aAs+iTGq + V
 21686 250 KRIGVPEEVSSVVCFLSPAASFITGQSVDV 280

Fig. 20

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
 Sheet 28 of 43
 28/43

ProDom Matches

ProdomId	Start	End	Description	Score
View Prodom 121622 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	29	82	p99.2 (1) YS05_CAEEL // HYPOTHETICAL 98.0 KD PROTEIN F56D1.5 IN CHROMOSOME II TRANSMEMBRANE	70
View Prodom 95301 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	35	82	p99.2 (1) O27957_ARCFU // SHIKIMATE 5-DEHYDROGENASE AROE HYPOTHETICAL PROTEIN	86
View Prodom 11 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	37	231	p99.2 (1078) ADH(34) GALE(20) FABG(13) // OXIDOREDUCTASE PROTEIN DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIS SYNTHASE ALCOHOL PUTATIVE	157
View Prodom 73753 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	237	286	p99.2 (1) P71079_BACSU // UNIDENTIFIED DEHYDROGENASE	84
View Prodom 77223 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	243	287	p99.2 (1) O07882_STAXY // GLUCOSE-1-DEHYDROGENASE	92
ProdomId	Start	End	Description	Score

View Prodom 11

>11 p99.2 (1078) ADH(34) GALE(20) FABG(13) // OXIDOREDUCTASE PROTEIN DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIA SYNTHASE ALCOHOL PUTATIVE Length = 269

Score = 157 (60.3 bits), Expect = 1.2e-09, P = 1.2e-09
 Identities = 64/213 (30%), Positives = 106/213 (49%)

Query: 51 KAIVKELLELGSNVVIASRKLERLKSAADELQANLPPTKQA---RVIPIQCNIRNEEVN 107
 K +V S AS+ E + + A + T QA V + C++ + E+V

Sbjct: 35 KVVVVSATSEEESTEASK--ESAMEVSKAVNAEVSATMQAVGVTVTKVTCVDADVEDVE 92

Query: 108 NLVKSTLDTF---GKINFLVNNGGQFLSP---AEHISSKG----WHAVALTNLTGTF 155
 LV++ ++ F GKI+ LVNN G ++P AE ++ + W V+E N+TGTf

Sbjct: 93 KLVETVVEEFSGIHGKIDVLVNNAG--VMAPKAVAESMTEETSDEEWEVIEVNVTGTF 150

Query: 156 YMCKAVYSSWMK-----EHGGSIVNI--IVPTKAGFP--LAVHSGAARAGVYNLTKS 203
 + +A + K G+IVN+ + + G P A +S A++A V + TKS

Sbjct: 151 NLTQAAALPAMKKFSDAAKKRFVGTIVNVASVAGSTMGSPGSQAAYS-ASKAAVESFTKS 209

Query: 204 LALE---WACSG--IRINCVAvgviySQTAVEN 231
 LA+E ++ S +R+N VAPG + + A+E+

Sbjct: 210 LAMELSPYSSASVAMVRVNAVAPGYVETD-ALES 241

Score = 103 (41.3 bits), Expect = 0.0021, Sum P(2) = 0.0021
 Identities = 32/100 (32%), Positives = 54/100 (54%)

Query: 37 GQVAIVTGGAA-TGIGKAIVKELLELGSNVVIASRKLERLKS--AADE-----LQAN 84
 G+ +VTGG+ +GIG AI ++L E G+ VV+ S E +S A+ E + A

Sbjct: 7 GKTVLVTGGSGFSGIGLAIARQLAEEGAKVVVVSATSEEESTEASKESAMEVSKAVNAE 66

Query: 85 LPPTKQA---RVIPIQCNIRNEEVNNLVKSTLDTFGKIN 121
 + T QA V + C++ + E+V LV++ ++ F I+

Sbjct: 67 VSATMQAVGVTVTKVTCVDADVEDVEKLVETVVEEFSGIH 106

Score = 37 (18.1 bits), Expect = 0.0021, Sum P(2) = 0.0021
 Identities = 9/23 (39%), Positives = 13/23 (56%)

Query: 205 ALEWACSGIRINCVAvgviySQ 227

ALE A +G+ + V PG + T

Sbjct: 238 ALESATNGLSVVTVRPGNVRVNT 260

Fig. 21A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 29 of 43

29/43

View Prodom 77223

>77223 p99.2 (1) 007882_STAXY // GLUCOSE-1-DEHYDROGENASE
Length = 67

Score = 92 (37.4 bits), Expect = 0.00031, P = 0.00031
Identities = 19/45 (42%), Positives = 29/45 (64%)

Query: 243 SFQKIPAKRIGVPEEVSSVVCFLSPAASFITGQSVDVGGRSLY 287
+ + IPAK IG +V++V FL S A +I G ++ VDGG + Y
Sbjct: 15 TLEMIPAKEIGFADQVANVARFLCSLDLADYIHGTTIYVDGGMTNY 59

View Prodom 95301

>95301 p99.2 (1) 027957_ARCFU // SHIKIMATE 5-DEHYDROGENASE
AROE HYPOTHETICAL
PROTEIN
Length = 108

Score = 86 (35.3 bits), Expect = 0.0014, P = 0.0014
Identities = 20/48 (41%), Positives = 31/48 (64%)

Query: 35 LQQQVAIVTGGATGIGKAIVKELLELGNSVVIASRKLERLKSAADELQ 82
L G+ A+V G A G GKA LL++GS V++A+R E+ + A + L+
Sbjct: 10 LGGKTALVVG-AGGAGKAAALALLDMGSTVIVANREEKGREAVEMLR 56

View Prodom 73753

>73753 p99.2 (1) P71079_BACSU // UNIDENTIFIED DEHYDROGENASE
Length = 60

Score = 84 (34.6 bits), Expect = 0.0023, P = 0.0023
Identities = 20/50 (40%), Positives = 29/50 (58%)

Query: 237 QSFFEGSFQKIPAKRIGVPEEVSSVVCFLSPAASFITGQSVDVGGRSL 286
+ E + Q PA R+ + + V FL+S A I GQ++ VDGGRSL
Sbjct: 9 EDLLEDARQNTPAGRMEIKDMVDTVEFLVSSKADMIRGQTIIVDGGRSL 58

View Prodom 121622

>121622 p99.2 (1) YS05 CAEEL // HYPOTHETICAL 98.0 KD PROTEIN F56D1.5 IN
CHROMOSOME II TRANSMEMBRANE
Length = 194

Score = 70 (29.7 bits), Expect = 7.6, P = 1.0
Identities = 20/57 (35%), Positives = 29/57 (50%)

Query: 29 YLAPG;;QGQV--AIVTGGATGIGKAIVKELLELG-SNVVIASRKLERLKSAADELQ 82
+ P L Q Q +V+GG GIGKA EL + G V+ R ++L S E++
Sbjct: 62 FYKPNLEQYQHRWTVVSGGTDGIGKAYTLEAKRGLRFVLLIGRNPKKLDSVKSEIE 118

Fig. 21B

30/43

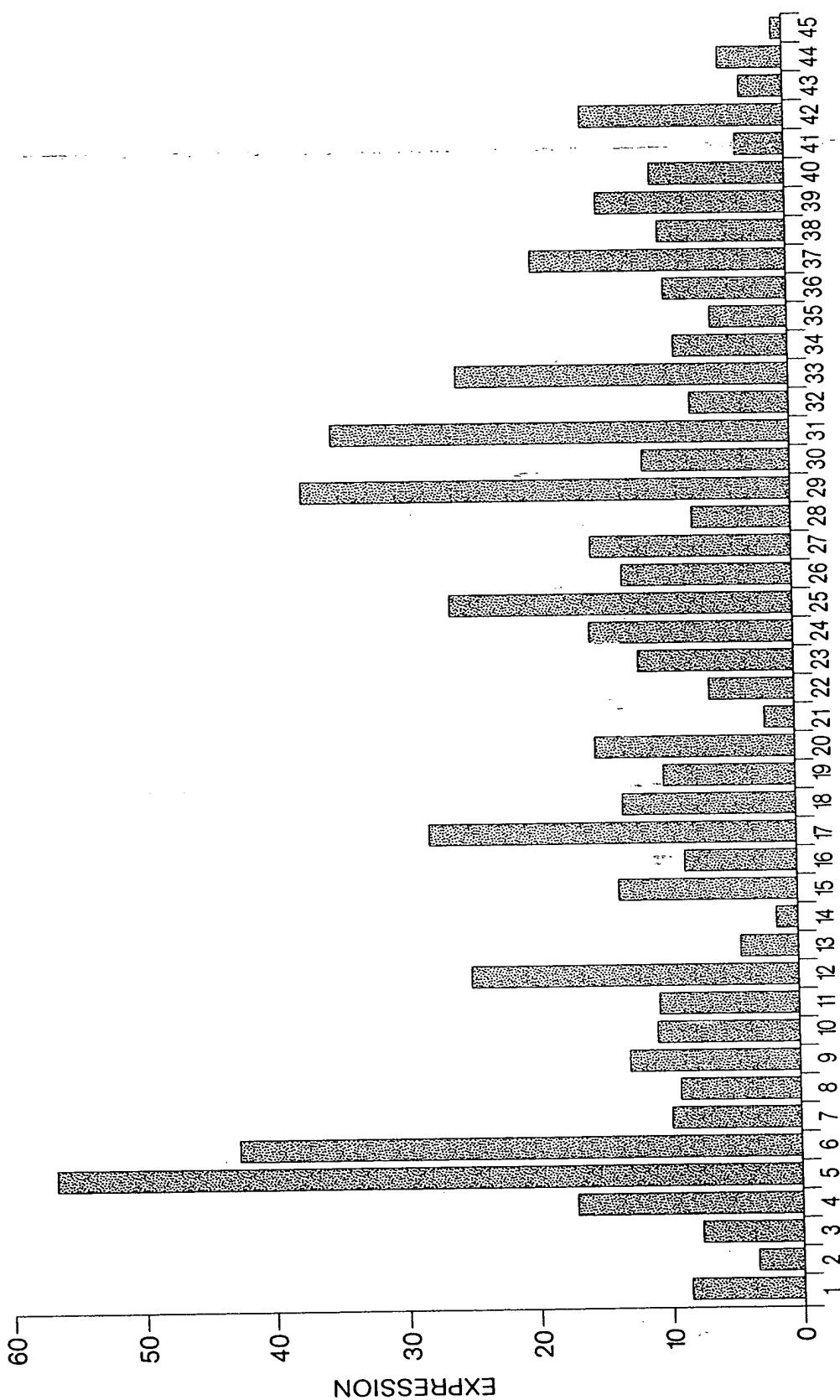


Fig. 22

31/43

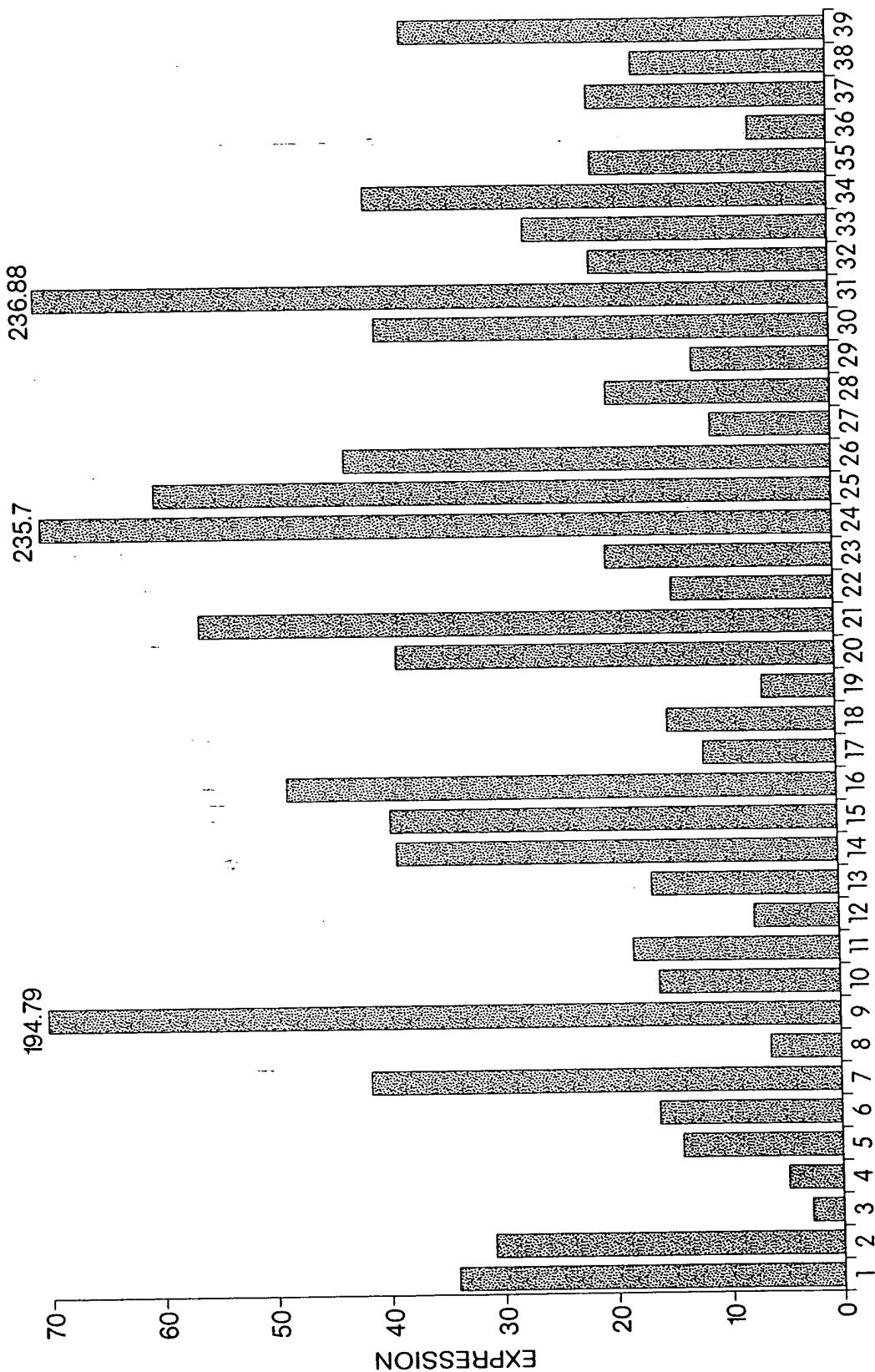


Fig. 23

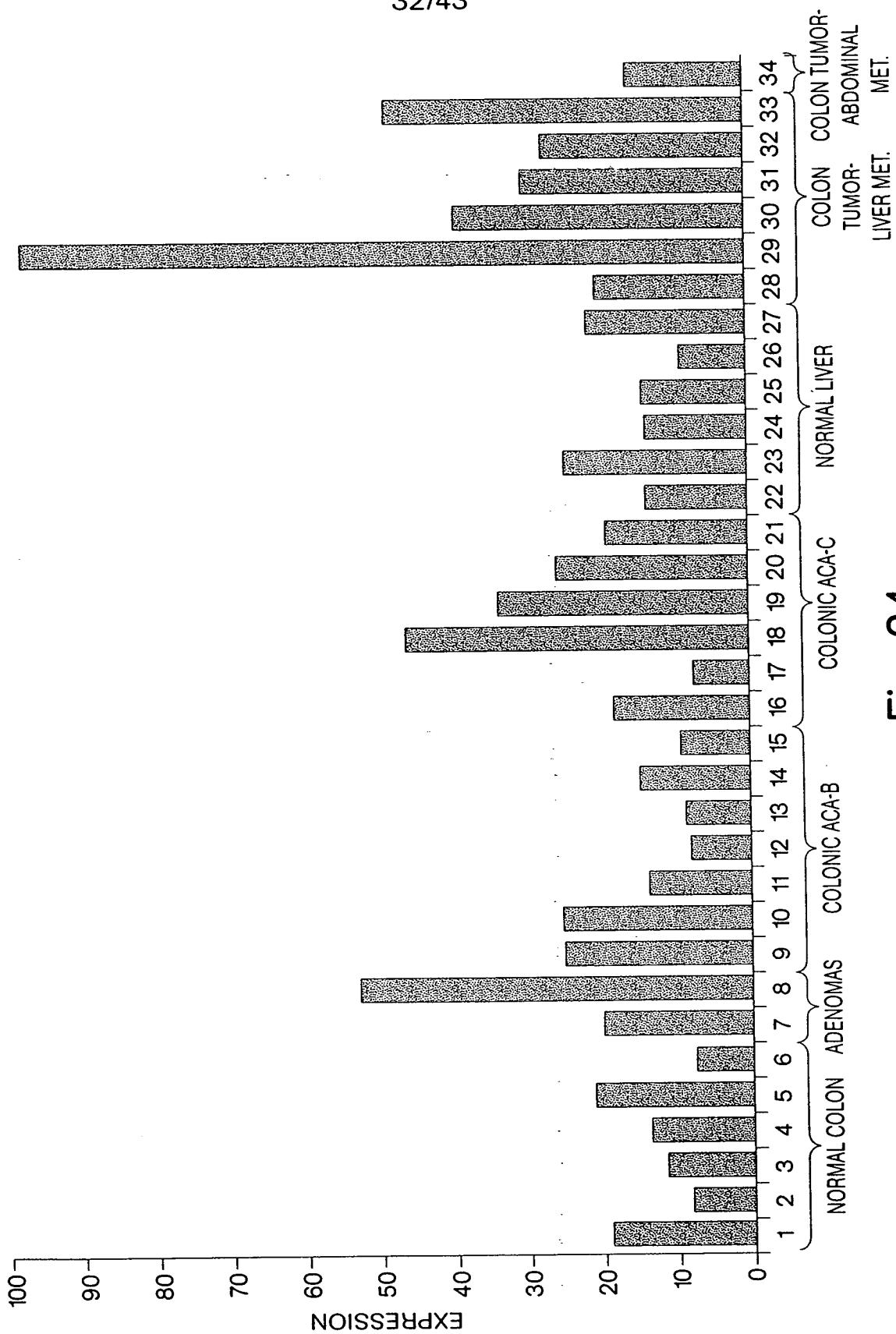


Fig. 24

33/43

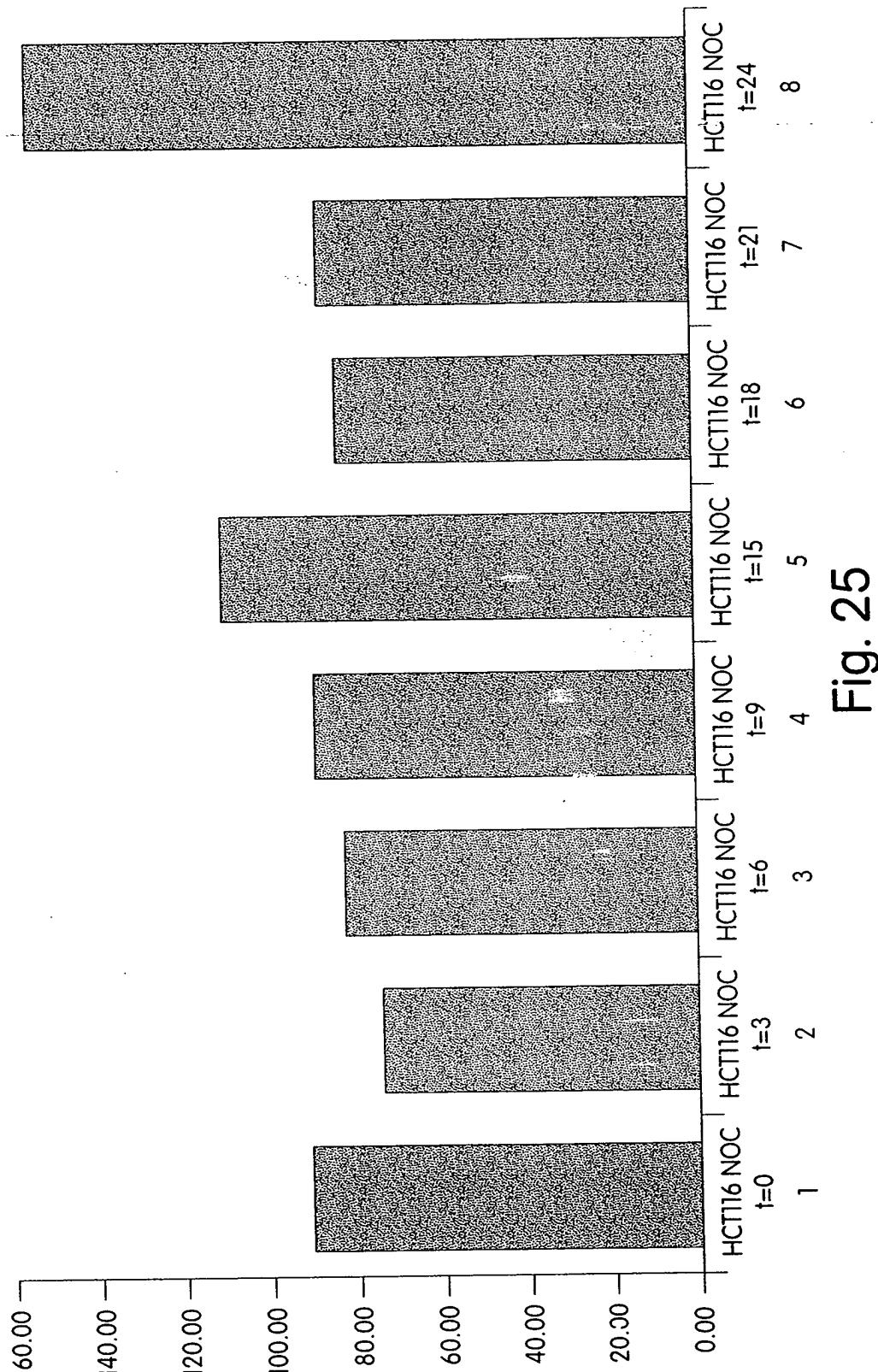


Fig. 25

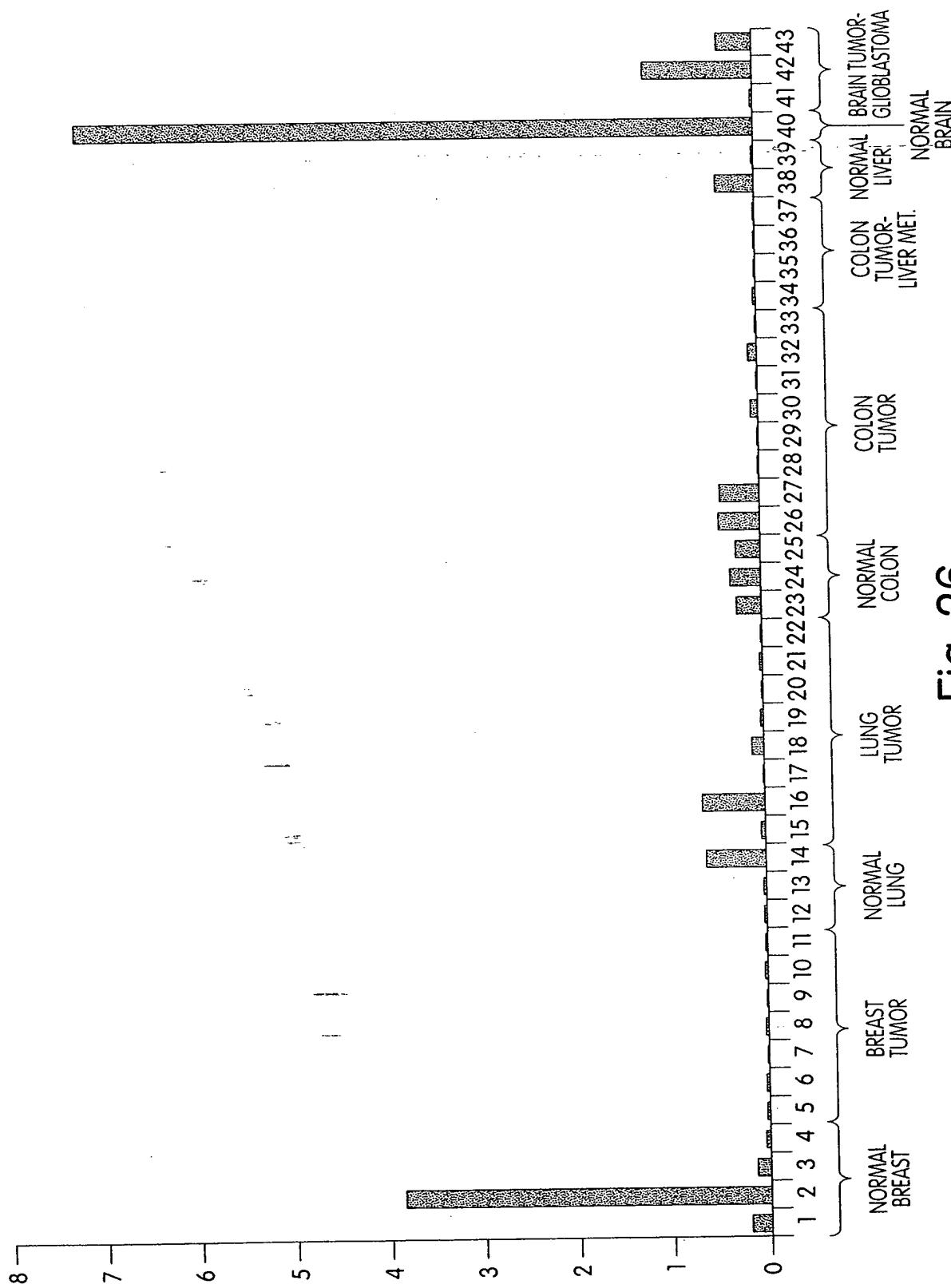


Fig. 26

35/43

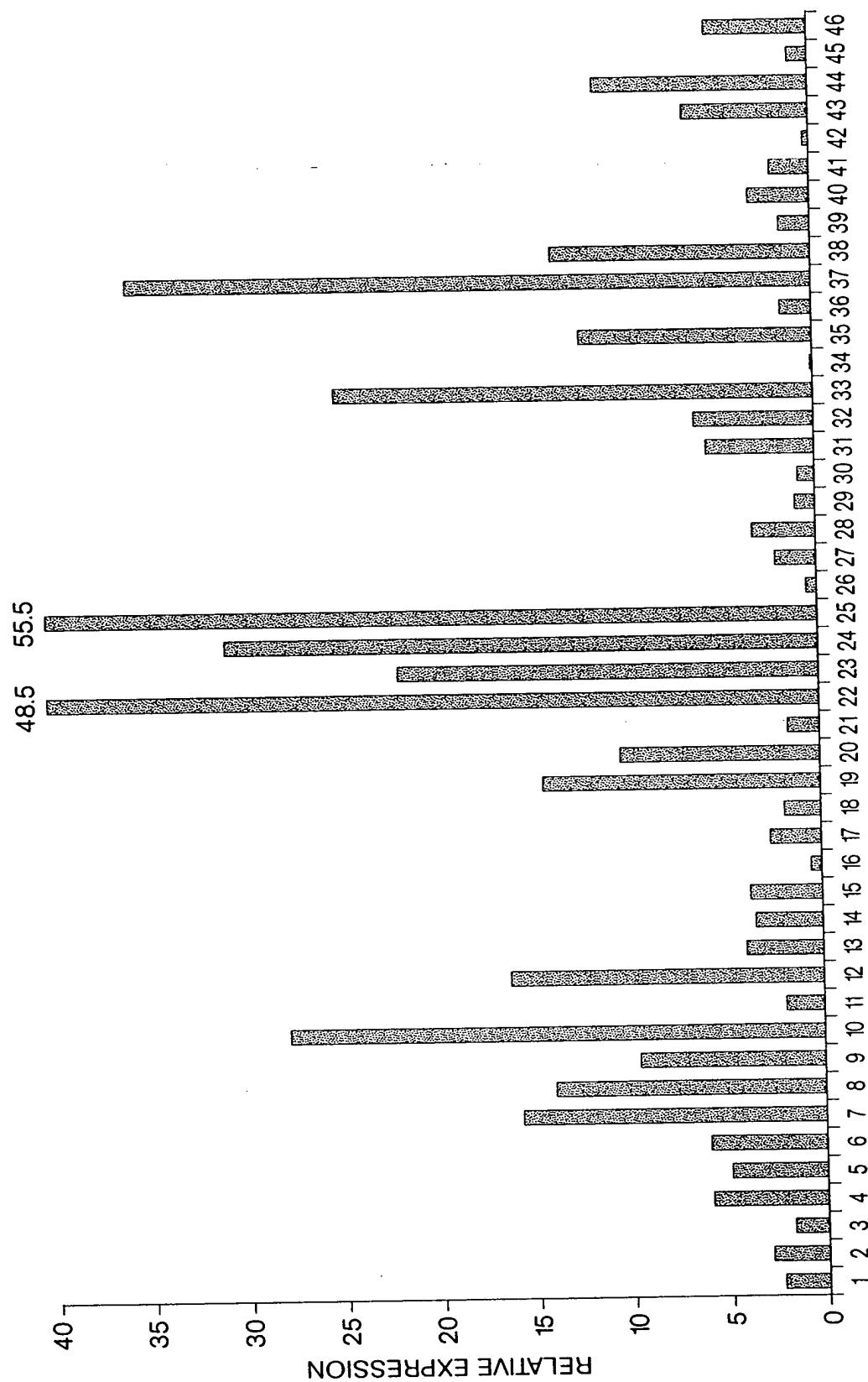


Fig. 27

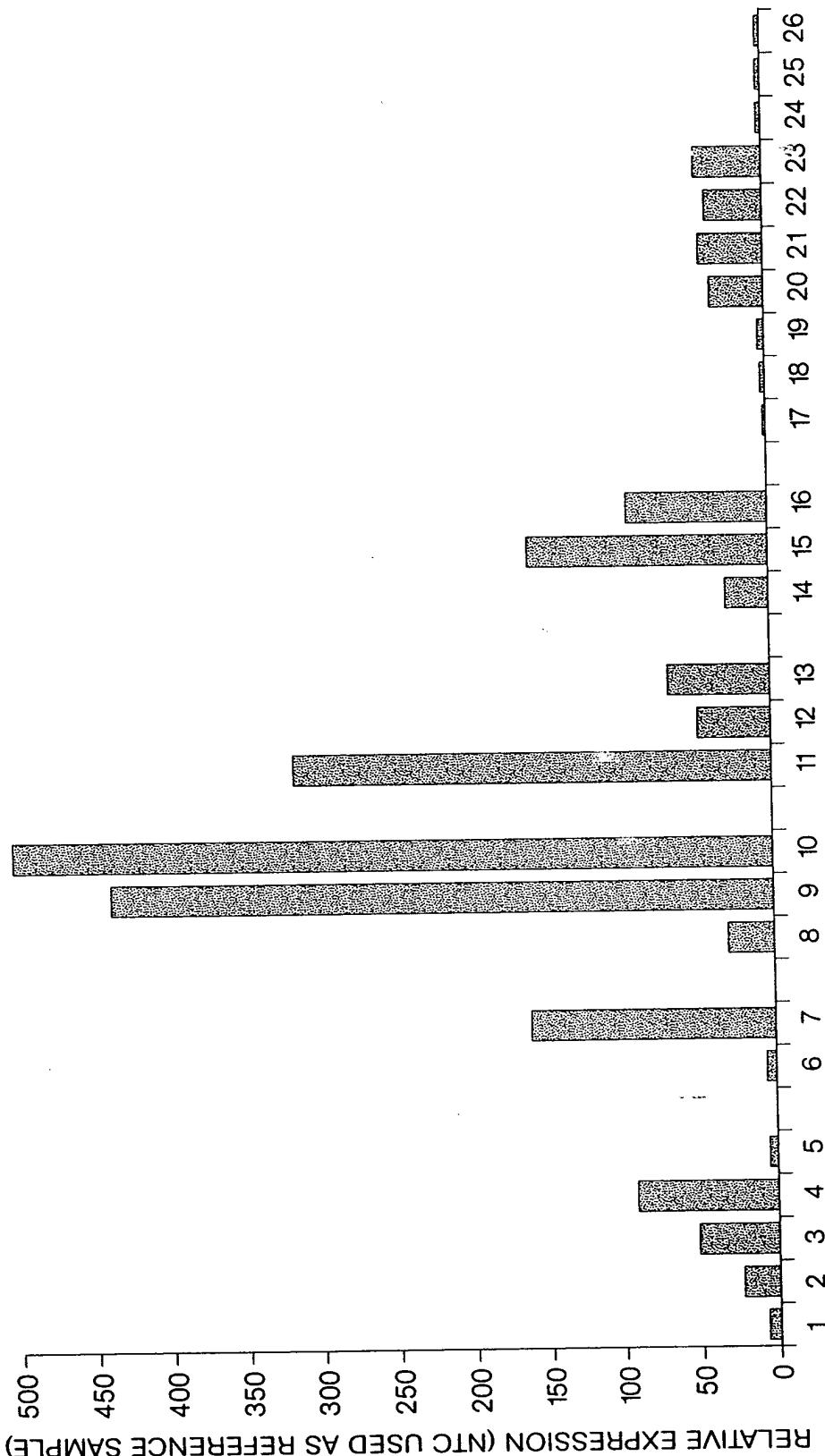


Fig. 28

37/43



Fig. 29

38/43

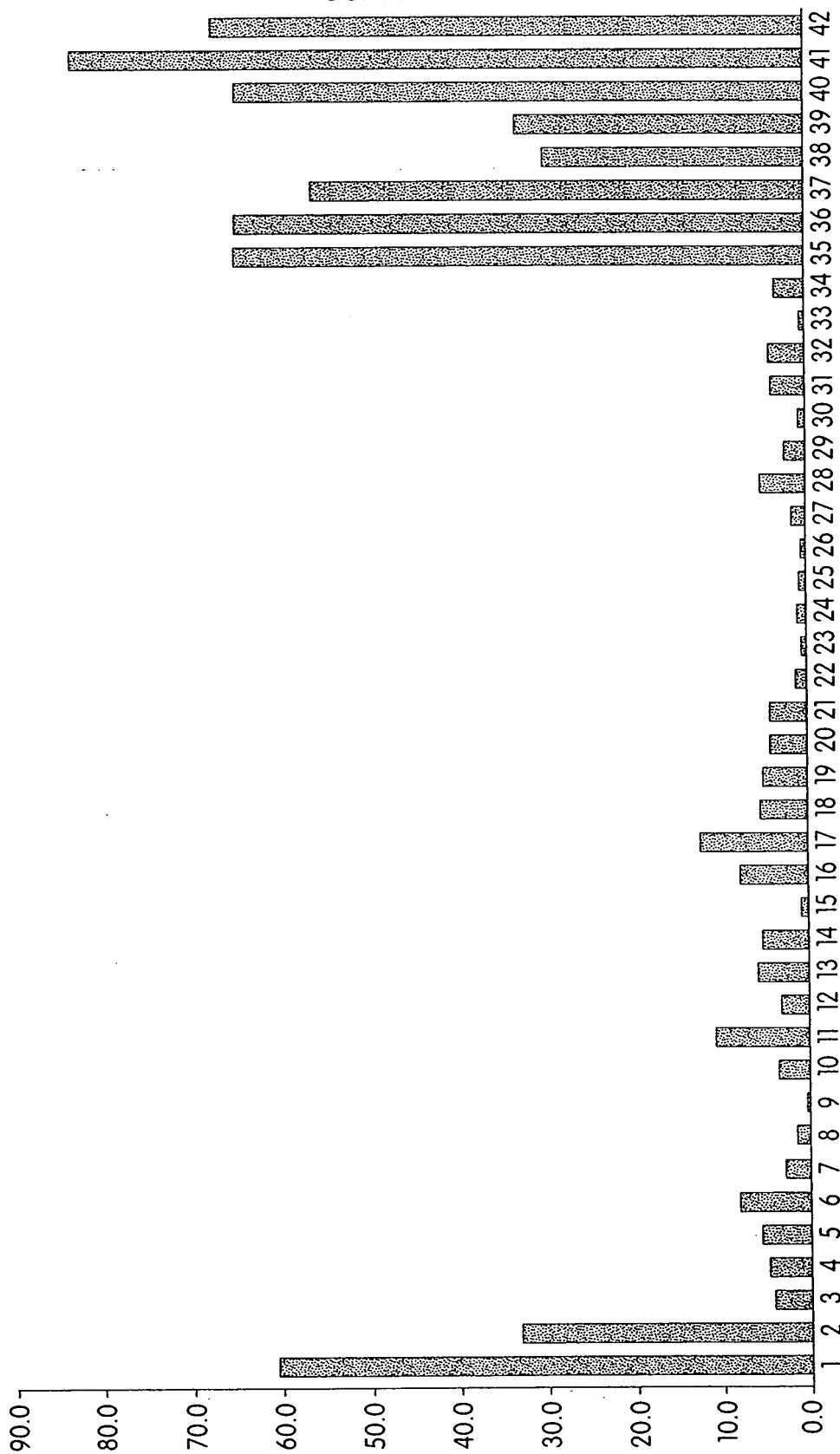


Fig. 30

39/43

GGAATGGATGCTGTTGGCTTAAACCTCCCCCTGCCCTGGGGTTGCAACCAGGGTCTCTG
CAAAGCCAATCCTTGTCATCCCGCTGTCCTGCAGAGCAAGATGGGGCTCATGGCTGTCC
TGATGCTACCCCTGCTGCTGGAAATCAGCGGCCTCCTCTTCATTACCAGGAGGCAT
CCAGGCTGTGGTCGAAGTCTGCCGTGCAGAACAAAGTGGTGGTCATCACAGATGCCATCT
CAGGACTGGAAAGGAGTGTGCTCGGTGTTCCATGCAGGTGGGCAAGGCTGGTGTGT
GTGGAAAGAACTGGGAGGGACTGGAGAGCCTCTATGCCACCTTGACCAGTGTGGCTGACC
CCAGCAAGACATTCACCCCAAGCTGGCCTCCTGGATCTCTCAGACATTAGCTGTGTTCA
AAGATGTGGCAAAGAGGTCTGGACTGCTACGGCTGTGGACATCCTCATCAACAATG
CCAGCGTAAAGTGAAGGGCCTGCCACAAGATTCCCTGGAGCTTGACAAAAAGATCA
TGGATGCCAACTACTCGGACCCATCACTTAACCAAAGTTCTGCTTCCAACATGATCT
CCAGGAGAACAGGCCAGATTGTGTTAGTGAACAAACATCCAAGCGAAGTTGGAATCCCGT
TCCGCACAGCTTATGCAGCCTCTAACCATGCCGTATGGCTTCTTGACTGCCCTCCGAG
CCGAGGTTGAGGAATACGATGTTGGTCAGCACCGTGAGGCCAACTTCATCCGCTCCT
ACCGTGCTTCCCTGAGCAAAGAAACTGGGAGACATCCATTGAAATTCTCTGCAGGA
AGCTAGCCTATGGCGTGCACCGGTGGAGGTGGCTGAGGAAGTGTGCGCACAGTACGGA
GGAAGAAGCAAGAGGTGTTCATGCCAACCCGGTCTTAAGGCTGCCGTGTTCATCCGCA
CCTTCTCCCTGAGTTCTCGCTGTGGTGGCCTGTGGGTGAAGGAGAAGCTCAATG
TCCCAGAAGAGGGTTAACCTCGTGGCAAAGGGTCACTCAAGGGAAATAAGGCTTCC
TAGAGAAAAAAAAAAAAAAA

Fig. 31A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 40 of 43

40/43

MGLMAVIMLPLLLGISGLLFIYQEASRLWSKSAVQNKVVVITDAISGLGKECARVFHAG
GARLVLCGKNWEGLESLYATLTSVADPSKTFTPKLVLLDLSDISCVQDVAKEVLDYCFCV
DILINNASVKVKGPAHKISLELDKKIMDANYFGPITLTKVLLPNMISRRTGQIVLVNNIQ
AKFGIIPFRAYAASKHAVMGFFDCLRAEVEEYDVVVSTVSPTFIRSYRASPEQRNWETSI
CKFFCRKLAYGVHPVEVAEEVMRTVRRKKQEVFMANPVPKAAVFIRFFFPEFFFAVVACG
VKEKLNVPEEG.

Fig. 31B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 41 of 43

41/43

GAP of: FrGcgManager_31_UFAHDJyG_ check: 516 from: 1 to: 936

M21481 ORF - Import - vector trimmed

to: FrGcgManager_31_VFA0zr_19 check: 2871 from: 1 to: 933

h21481 ORF - Import - vector trimmed

Symbol comparison table: /ddm_local/gcg/gcg_9.1/gcgcore/data/rundata/nwsgapdna.cmp

CompCheck: 8760

Gap Weight: 12 Average Match: 10.000
Length Weight: 4 Average Mismatch: 0.000

Quality: 8220 Length: 936
Ratio: 8.810 Gaps: 0

Percent Similarity: 88.103 Percent Identity: 88.103

Match display thresholds for the alignment(s):

| = IDENTITY
: = 5
. = 1

FrGcgManager_31_UFAHDJyG_ x FrGcgManager_31_VFA0zr_19 ..

1 ATGGGGCTCATGGCTGTCCTGATGCTACCCCTGCTGCTGCTGGGAATCAG 50
1 atgggagtcatggccatgctgatgtcccccgtctgtgtggaaatcag 50
51 CGGCCTCCTCTTCATTACCAGGAGGCATCCAGGCTGTGGTCGAAGTCTG 100
51 cggcctcctttcatttaccaagagggtgtccaggctgtggtaaaagtca 100
101 CCGTGCAGAACAAAGTGGTGGTCATCACAGATGCCATCTCAGGACTGGGA 150
101 ctgtgcagaacaaaagtggtggtgatcacccatgtcaggactgggc 150
151 AAGGAGTGTGCTCGGGTGTCCATGCAGGTGGGCAAGGCTGGTGCTGTG 200
151 aaggagtgctcggtgttccacacaggctggggcaaggctggtgctgt 200
201 TGGAAAGAACTGGGAGGGACTGGAGAGCCTCTATGCCACCTTGACCAAGTG 250
201 tggaaagaactgggagaggctagagaaccttatatgatgcctgatcagcg 250
251 TGGCTGACCCCAGCAAGACATTCAACCCCCAAGCTGGTCCTGGATCTC 300
251 tggctgaccccagcaagacattcacccaaagctggtcctgtggacctc 300

Fig. 32A

301 TCAGACATTAGCTGTGTTCAAGATGTGGCAAAGAGGTCTGGACTGCTA 350
301 tcagacatcagctgtgtccagatgtggaaaagaagtctggattgcta 350
351 CGGCTGTGTGGACATCCTCATCAACAATGCCAGCGTAAAGTGAAGGGC 400
351 tggctgtgtggacatcctcatcaacaatgccagtgtgaaggtgaaggggc 400
401 CTGCCACACAAGATTCCCTGGAGCTTGACAAAAGATCATGGATGCCAAC 450
401 ctgcccataagattctctggagctcgacaaaaagatcatggatgcaat 450
451 TACTTCGGACCCATCACTTAACCAAAGTTCTGCTTCCCACATGATCTC 500
451 tactttggccccatcacattgacgaaagccctgcttcccaacatgatctc 500
501 CAGGAGAACAGGCCAGATTGTGTTAGTGAACACATCCAAGCGAAGTTG 550
501 ccggagaacaggccaaatcgtgttagtgaataatatccaaggaaagttg 550
551 GAATCCCGTCCGCACAGCTTATGCAGCCTCTAACGCATGCCGTATGGGC 600
551 gaatcccggtccgtacgacttacgctgcctccaaggcacgcagccctggc 600
601 TTCTTGACTGCCTCCGAGCGAGGTTGACCAATACGATGTTGTGGTCAG 650
601 ttcttgactgcctccgagccgaaatggagatcgatgttgcattcag 650
651 CACCGTGAGCCAACTTCATCCGCTCCTACCGTCTTCCCCTGAGCAA 700
651 caccgtgagcccgactttcatccggctgtaccacgttatccagagcaag 700
701 GAAACTGGGAGACATCCATTGTAAATTCTTCTGCAGGAAGCTAGCCTAT 750
701 gaaactgggaaagcttcattggaaattctttcaggaagctgactac 750
751 GGC GTGCACCCGGTGGAGGTGGCTGAGGAAGTGATGCGCACAGTACGGAG 800
751 ggcgtgcacccacttagaggtggcgaggaggtatgcgcaccgtgcggag 800
801 GAAGAAGCAAGAGGTGTTCATGGCCAACCCGGTCTAACGGCTGCCGT 850
801 gaagaagcaagagggtttatggccaacccatccccaaaggccgcgtgt 850
851 TCATCCGCACCTTCTCCCTGAGTTCTTCTCGCTGTGGTGGCCTGTGGG 900
851 acgtccgcacccatccggatgttttcgcgtggccctgtggg 900
901 GTGAAGGAGAAGCTCAATGTCCCAGAAGAGGGTTAA 936
901 gtgaaggagaagctcaatgtcccgaggagg... 933

Fig. 32B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 43 of 43
43/43

GAP of: FrGcgManager_32_ZFA004eiD check: 657 from: 1 to: 311

m21481 aa - Import - complete

to: FrGcgManager_32_AGAjaPna_ check: 9949 from: 1 to: 311

h21481 aa - Import - complete

Symbol comparison table: /prod/ddm/seqanal/BLAST/matrix/aa/BLOSUM62

CompCheck: 1102

Matrix made by matblas from blosum62.iij

Gap Weight: 12 Average Match: 2.778
Length Weight: 4 Average Mismatch: -2.248

Quality: 1467 Length: 311
Ratio: 4.717 Gaps: 0

Percent Similarity: 92.926 Percent Identity: 91.318

Match display thresholds for the alignment(s):

| = IDENTITY
: = 2
. = 1

FrGcgManager_32_ZFA004eiD x FrGcgManager_32_AGAjaPna_ ..

1 MGLMAVLMILLLLLGISGLLFIYQEASRLWSKSAVQNKVVVITDAISGLG 50

1 MGVMAMLMLPLLLLGISGLLFIYQEVSRLWSKSAVQNKVVVITDAISGLG 50

51 KECARVFHAGGARLVLCGKNWEGLESLYATLTSVADPSKTFTPKLVLLDL 100

51 KECARVFHTGGARLVLCGKNWERLENLYDALISVADPSKTFTPKLVLLDL 100

101 SDISCVQDVAKEVLD CYGCV DILINNA SVVKV GPAHK I SLELDKKIM DAN 150

101 SDISCVPDVAKEVLD CYGCV DILINNA SVVKV GPAHK I SLELDKKIM DAN 150

151 YFGPITLTkvllPNMISRRTGQIVLVNNIQAKFGIPFR TAYAASKHA VMG 200

151 YFGPITLTkALLPNMISRRTGQIVLVNNIQGKFGIPFR TTYAASKHA ALG 200

201 FFDCLRAEVEEYDV VSTVSPTFIRSYRASPEQRNWETSICKFFCRK LAY 250

201 FFDCLRAEVEEYDV VSTVSPTFIRSYHV YPEQGNWEASIWKFFF RKL TY 250

251 GVHPVEVAEEVMRTVRRKKQEVFMANPVPKA AVFIRTFFPEFFF AVVACG 300

251 GVHPVEVAEEVMRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFF AVVACG 300

301 VKEKLNVPEEG 311

301 VKEKLNVPEEG 311

Fig. 33